ANÁLISIS DE FACTORES QUE INFLUYEN EN EL SECUESTRO DE CARBONO EN LOS SUELOS Y SU RELACIÓN CON LA COMPOSICIÓN MOLECULAR DE LA MATERIA ORGÁNICA

ANALYSIS OF FACTORS INVOLVED IN SOIL CARBON SEQUESTRATION AND THEIR RELATIONSHIP WITH THE MOLECULAR COMPOSITION OF ORGANIC MATTER

TESIS DOCTORAL / DOCTORAL THESIS

Marco Antonio Jiménez González

Madrid, 2019

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Geología y Geoquímica

Análisis de factores que influyen en el secuestro de carbono en los suelos y su relación con la composición molecular de la materia orgánica

TESIS DOCTORAL

Marco Antonio Jiménez González

Madrid, 2019

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Geología y Geoquímica

Análisis de factores que influyen en el secuestro de carbono en los suelos y su relación con la composición molecular de la materia orgánica

MEMORIA presentada por

Marco Antonio Jiménez González

para optar al grado de

Doctor con Mención Internacional

Directores: Dra. Ana María Álvarez González
Dr. Gonzalo Almendros Martín

Madrid, 2019

La **Dra. Ana María Álvarez González**, Profesora Titular de la Universidad Autónoma de Madrid del Departamento de Geología y Geoquímica y el **Dr. Gonzalo Almendros Martín**, Profesor de Investigación del Museo Nacional de Ciencias Naturales de Madrid del Consejo Superior de Investigaciones Científicas,

HACEN CONSTAR

Que la presente memoria titulada "Análisis de factores que influyen en el secuestro de carbono en los suelos y su relación con la composición molecular de la materia orgánica", presentada por el doctorando Marco Antonio Jiménez González ha sido realizada bajo su dirección, en el Departamento de Biogeoquímica y Ecología Microbiana del Museo Nacional de Ciencias Naturales de Madrid, reuniendo los requisitos de originalidad y rigor científico necesarios para ser presentada como Tesis Doctoral.

Y para que conste a efectos oportunos, dan el visto bueno para su presentación ante la Comisión de Doctorado de la Universidad Autónoma de Madrid

En Madrid, a 2 de septiembre de 2019,

Dra. Ana María Álvarez González Dr. Gonzalo Almendros Martín

La realización del presente trabajo ha sido posible gracias al apoyo económico recibido a través del proyecto CGL2013-43845-P y a la concesión de una beca de Formación de Personal Investigador (FPI, BES-2014-069238) concedidos por el Ministerio de Economía y Competitividad

Agradecimientos

Como no agradecer este trabajo a mis dos directores, la Dra. Ana María Álvarez González y el Dr. Gonzalo Almendros Martín, sin ellos este trabajo no hubiera sido posible. Guiándome en estos cuatro años, con todos esos buenos momentos en los que he aprendido muchísimo. Han sido cuatro años muy duros e importantes de mi vida en los que he tenido la suerte de teneros a vosotros, de nuevo mil gracias por estar a mi lado.

Agradecer a los profesores Raimundo, Victoria, Pilar, María José y Zulimar por todos esos buenos momentos en los laboratorios en la Universidad Autónoma de Madrid y en las salidas de campo con los alumnos, en las cuales he disfrutado y aprendido mucho. Por supuesto, también a Elena con esa increíble energía que siempre tiene en el laboratorio y a Amalia por toda su ayuda.

Agradecer a la Dra. María del Carmen Lobo, por facilitarme la determinación de propiedades hídricas de los suelos en sus laboratorios del Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario (IMIDRA).

Dar las gracias al profesor Patrick Hatcher y todo su grupo, por acogerme en su laboratorio durante mi estancia en Estados Unidos en Old Dominion University y ayudarme en el uso de nuevas técnicas como la FTICR-MS y en la interpretación de sus resultados en esta Tesis doctoral. A Gail por esos inolvidables momentos durante mi estancia y por acogerme con esa increíble amabilidad.

No olvidar a todos los colegas y amigos sevillanos del Instituto de Recursos Naturales y Agrobiología de Sevilla, los doctores: José Antonio, Heike, José María, Ana y como no, al Prof. Francisco Javier González-Vila, por esos fantásticos momentos que hemos pasado y los que nos quedan por pasar. Gracias por haberme ayudado en la utilización equipos para el análisis e interpretación de los resultados obtenidos de las muestras empleadas en este trabajo, mil gracias a todos.

A mi gran amigo y compañero el Dr. Nicasio T. Jiménez-Morillo: solo espero que algún día podamos trabajar juntos otra vez, codo con codo. Aún recuerdo como si fuera ayer, el día que ambos decidimos hacer juntos nuestro trabajo sobre

suelos afectados por incendios con los Profesores Antonio y Lorena, cuando aún éramos estudiantes de Química. Un enorme agradecimiento a estos dos profesores con los que empezó todo, los que despertaron mi interés por los suelos, mil gracias a vosotros y a todo el grupo MED_Soil (Gael, Arturo, Jorge, Angel, Eli, etc.).

No olvidar esa fantástica estancia en University of New South Wales en Sídney (Australia) con la Dra. Miriam Muñoz-Rojas, mil gracias por esa increíble estancia en esa tierra tan especial en la que he aprendido muchísimo y me he sentido como en casa gracias a ti y tu familia. Agradecer también al Dr. Mark Ooi su amabilidad durante toda mi estancia.

Dar las gracias a mis dos compañeros de piso, Javi y Alberto, por estos cuatro años de perfecta convivencia en Madrid, dos grandes amigos que me llevo de estos maravillosos años. Sinceramente pienso que se han convertido en expertos de materia orgánica en suelos, ya que me han visto en esos momentos de stress total con mi trabajo y se han tragado todas esas explicaciones que les daba sobre el carbono en el suelo.

No olvidar a mis amigos de toda la vida en mi tierra natal Mairena, ese grupo de amigos del que uno puede presumir, amigos de verdad. A pesar de las distancias siempre están ahí, gracias a todos vosotros.

Por supuesto a todo el equipo de desayuno, Luis, Elena, Mercedes, Gema, Carmen y María, gracias a todos vosotros por todas esas risas durante el desayuno.

Por último, agradecer a mis padres Catalina y Antonio por estar siempre ahí, a pesar de las distancias, siempre encontrábamos tiempo para esas innumerables llamadas por Skype. Ellos que siempre han estado junto a mí, en ese largo camino académico, apoyándome y ayudando en todo momento, gran parte de todo esto lleva vuestro sello. A mi hermano Rubén, por todas esas visitas a Madrid, por esas grandes risas y enorme alegría en todo momento, por esa gran conexión y complicidad de hermanos que tenemos, mil gracias. Gracias a toda mi familia, a los que están y a los que a pesar de que nos dejaron sé que me siguen observando desde donde quiera que estén, este trabajo tiene parte de vosotros. Con especial cariño a mi abuela Cristina, siempre con esa sonrisa en las videollamadas cuando me encontraba en EEUU o Australia.

A mis padres,

Cati y Antonio

Acrónimos / Acronyms

ACL: Longitud media de la cadena / Average chain length

AIC: Criterio de información de Akaike / Akaike's information criterion

AI_{mod}: Índice de aromaticidad modificado / Modified aromaticity index

CEC: Capacidad de intercambio catiónico / Cation exchange capacity

CP: Polarización cruzada / Cross polarization

CPI: Índice de preferencia de carbono / Carbon preference index

DBE: Dobles enlaces equivalentes / Double bond equivalents

DHPQ: 4,9-dihidroxiperilen-3,10-quinona / 4,9-dihydroxyperylene-3,10-quinone

EC: Conductividad eléctrica / Electrical conductivity

ESI: ionización de electrospray / Electrospray ionization

FA: Ácido fúlvico / Fulvic acid

FFA: Ácido fúlvico libre / Free fulvic acid

FOM: Materia órganica libre / Free organic matter

FTICR-MS: Espectrometría de masas de resonancia ciclotrónica de ión de transformada de Fourier / Fourier transform ion cyclotron resonance mass spectrometry

H': Índice de diversidad de Shannon-Wiener / Shannon-Wiener diversity index

HA: Ácido húmico / Humic acid

IR: Espectroscopía infrarroja / Infrared spectroscopy

LV: Variables latentes / Latent variables

MAS: Rotación en el ángulo mágico / Magic-angle spinning

MDS: Escalado multidimensional / Multidimensional scaling

NMR: Espectroscopía de resonancia magnética nuclear / Nuclear magnetic resonance spectroscopy

PCA: Ánalisis de componentes principales / Principal components analysis

PLS: Regresión de mínimos cuadrados parciales / Partial least square regression

PRESS: Suma de los cuadrados de error de predicción / Average prediction sum of squares

Py-GC/MS: Pirólisis acoplada a cromatografía de gases espectrometría de masas / Pyrolysis-gas chromatography-mass spectrometry

Q: Índice de Emberger / Emberger's index

RMSE: Error cuadrático medio / Root mean square error

RT: Tiempo de retención / Retention time

SOC: Carbono órganico del suelo / Soil organic carbon

THE: Extracto húmico total / Total humic extract

VIP: Importancia de la variable en la proyección / Variable Importance for Projection

WHC: Capacidad de retención de agua / Water holding capacity

ÍNDICE

RESUMEN

El avance de la desertificación, muy acusado en el área del Mediterráneo, junto con la emisión a la atmósfera de gases de efecto invernadero, en especial el CO_2, es objeto de creciente atención por parte de la comunidad científica. El reconocimiento de los factores que intervienen en estos procesos, pero sobre todo la aplicación de medidas para retrasar su avance, está siendo objeto de análisis desde diversos ámbitos de la investigación. En este contexto, la conservación de los suelos presenta un especial interés, por cuanto el contenido de materia orgánica del suelo (SOM) y su estabilidad frente a la degradación son factores determinantes. En particular, los procesos biogeoquímicos de estabilización del carbono orgánico en los suelos (SOC) están siendo estudiados. Diversas investigaciones se centran en la estabilización de la SOM a consecuencia de interacciones órgano-minerales, mientras que otras se basan en establecer relaciones entre la estructura molecular de la SOM y su resistencia a la biodegradación. En este trabajo de investigación se abordó la caracterización molecular de la materia orgánica en diferentes tipos de suelo para establecer su relación con el potencial de almacenamiento de carbono de los correspondientes suelos, así como con los factores que influyen en su calidad. Para ello, se han seleccionado 35 suelos con alta variabilidad en su contenido en carbono orgánico (17–157 $g \cdot kg^{-1}$) y se ha llevado a cabo una caracterización detallada de la materia orgánica empleando técnicas destructivas y no destructivas que incluyen la pirólisis analítica (Py-GC/MS), la resonancia magnética nuclear (NMR) de ^{13}C en estado sólido, la espectroscopía infrarroja (IR), la espectroscopía visible y la espectrometría de masas de resonancia ciclotrónica de ión (FTICR-MS). En particular, se ha centrado la atención en determinados tipos de compuestos biomarcadores que pudieran actuar como indicadores de la incidencia de los procesos biogeoquímicos del suelo. Se han analizado con detalle las poblaciones de alcanos y de metoxifenoles, cuya composición molecular permite distinguir entre productos de síntesis microbiana o de transformacion del material vegetal.

En este estudio se examinó la utilidad del índice de diversidad de Shannon (*H'*) calculado a partir de las abundancias de alcanos obtenidos por pirólisis

(C_9–C_{31}) para la evaluación del potencial de almacenamiento y calidad de la materia orgánica de los suelos. Una serie de tratamientos multivariantes de datos mostraron una relación significativa entre la diversidad *H′* de los alcanos y la concentración de SOC. En particular, se encontró una relación significativa entre los niveles de carbono y el porcentaje de alcanos de cadena larga, mientras que el porcentaje de alcanos de cadena corta se correlacionó con descriptores específicos de la calidad de la SOM. Finalmente, la regresión por mínimos cuadrados parciales (PLS) permitió predecir el contenido de carbono utilizando exclusivamente la información proporcionada por los patrones de alcanos.

En un estudio paralelo, fueron comparados los diferentes tipos de metoxifenoles obtenidos a partir de muestras de suelo completo, aplicando pirolisis analítica y utilizando el índice de diversidad de Shannon para describir la complejidad de sus patrones pirolíticos. Se aplicaron una serie de análisis estadísticos exploratorios (regresión lineal, PLS, escalado multidimensional (MDS), etc) para analizar las relaciones existentes entre los productos de pirólisis y un conjunto de características químicas y espectroscópicas de la SOM así como el contenido en carbono de los suelos. Estos tratamientos mostraron que existen correlaciones significativas entre la progresiva diversidad molecular del conjunto de metoxifenoles y la concentración de carbono orgánico. El hecho de que la diversidad en la firma fenólica proporcione información acerca del potencial de almacenamiento de carbono en los suelos se puede interpretar como un efecto de la progresiva complejidad estructural de las macromoléculas vegetales modificadas por los microorganismos, que se vuelven difícilmente reconocibles por las enzimas. Desde un punto de vista cuantitativo, los modelos de regresión por PLS, utilizando como descriptores exclusivamente las abundancias totales de los 12 metoxifenoles mayoritarios, fueron especialmente eficaces para pronosticar el almacenamiento de carbono en el suelo.

Una vez estudiada la información proporcionada por los productos de pirólisis de la SOM en escenarios de distinta actividad de almacenamiento de carbono, se orientó el estudio hacia la composición de la fracción tradicionalmente considerada más representativa de la SOM, los ácidos húmicos (HA), que corresponden a productos coloidales de transformación avanzada de la

biomasa vegetal y microbiana. En una primera parte del estudio de esta fracción se realizó una caracterización de los HAs empleando espectroscopía visible, IR y NMR. Un estudio por PLS utilizando como descriptores las intensidades de los puntos de los espectros digitales de IR (4000–400 cm^{-1}) mostró que existe una relación entre el patrón espectral de IR y el contenido de SOC. Este fue también el caso del índice E4 (relacionado con el progreso de humificación, y basado en la densidad óptica de los HA a 465 nm). La aplicación del análisis de componentes principales (PCA) y el MDS sugirieron que las bandas asignadas a los grupos carbonilo y amida eran características en los HA de los suelos con bajo contenido de C, mientras que los espectros de los HA en los suelos con altos niveles de C mostraron un típico patrón de bandas de lignina, indicando acumulación de residuos vegetales menos transformados. Los perfiles espectrales se analizaron en detalle mediante tratamientos digitales que incluyen sustracción ponderada de espectros IR obtenidos al promediar los espectros de los HAs extraídos de los suelos clasificados en los cuartiles superior e inferior de acuerdo a la distribución del SOC. Tras calcular el nivel de significación estadístico de las diferencias, los resultados mostraron que la composición molecular de los HAs difería significativamente tanto en función del SOC como de la E4, de forma que los picos correspondientes a grupos aromáticos, carboxilo y amida presentaban mayor intensidad en los HAs de suelos con bajo contenido de SOC, mientras que los correspondientes a estructuras derivadas de lignina estaban más marcados en los espectros de los HAs de suelos con alto contenido de SOC. En una segunda parte del estudio de los HAs, se realizó su caracterización por medio de FTICR-MS. La aplicación de la PLS a los datos de FTICR-MS mostró un gran potencial en la identificación de los componentes moleculares de la materia orgánica, lo que permitió reconocer que variaban en función de los niveles de SOC en los diferentes entornos. Se obtuvo un modelo de pronóstico del SOC por medio de PLS utilizando como descriptores los 131 compuestos comunes (presentes en todos los HAs) detectados por FTICR-MS. Con objeto de identificar los compuestos con mayor valor como indicadores de los niveles de carbono, la importancia de las variables para la predicción (VIP) se representó en el espacio definido por las relaciones atómicas de los correspondientes compuestos utilizando diagramas de van Krevelen. Los resultados indicaron que existe una relación significativa entre

la composición molecular de los HAs y los niveles de C almacenado en el suelo: los HA en los suelos con altos niveles de C orgánico presentaban proporciones significativamente más altas ($P < 0.1$) de lípidos insaturados y lignina. Por otra parte, los niveles bajos de SOC se encontraron asociados a proporciones comparativamente más altas de compuestos alifáticos saturados.

La influencia variable de los factores ambientales que afectan el secuestro de carbono en los suelos, así como la calidad de la SOM almacenada en los mismos es poco conocida, si bien tiene gran trascendencia en las tasas de liberación de CO_2 a la atmósfera y el avance de la desertificación. El objetivo es estudiar la composición molecular de la SOM para evaluar la medida en que dichos factores pueden afectar sus características, además de los niveles de SOC. Los 193 compuestos mayoritarios liberados por pirólisis a partir de muestras de suelo completo se incluyeron en un estudio estadístico basado en análisis discriminante para evaluar el impacto de los clásicos factores de formación de los suelos (clima, vegetación y sustrato geológico) en el contenido total y la composición de la SOM. Se construyeron diagramas de van Krevelen para facilitar el reconocimiento de patrones característicos en la composición de SOM dependiendo de los factores de formación del suelo, que mostraron una influencia decreciente en el orden: clima > vegetación > sustrato geológico. El hecho de que la composición molecular de la SOM varíe sistemáticamente de acuerdo con factores ambientales, unido a los niveles totales de SOM significativamente diferentes dependiendo de los escenarios estudiados, sugiere un control ambiental tanto cualitativo como cuantitativo del secuestro de C en el suelo.

El progresivo avance de la desertificación de los suelos, asociado al agotamiento de la SOM en los ecosistemas mediterráneos, los expone a severos riesgos de erosión y desertificación. Por consiguiente, se llevó a cabo un estudio orientado a valorar el efecto de parámetros climáticos en la composición y contenido total de SOM. La composición molecular estudiada por Py-GC/MS se analizó frente a los valores del clásico cociente pluviotérmico de Emberger (Q). El empleo de la PLS, utilizando los compuestos de pirólisis como descriptores, permitió predecir ($P < 0.05$) el índice Q, y posteriormente identificar, a modo de *proxies* moleculares, los

compuestos que responden más sistemáticamente a la variabilidad climática. Para ilustrar las diferencias, los conjuntos de compuestos pirolíticos de suelos desarrollados bajo condiciones climáticas extremas se compararon utilizando diagramas de van Krevelen. Los resultados muestran que la estructura molecular de la SOM retiene información ambiental sobre el cociente pluviotérmico *Q*, que se refleja principalmente en la abundancia total de metoxifenoles y compuestos de tipo alquilbenceno. Esto sugiere que la evolución de la SOM se estabiliza en etapas más o menos avanzadas de transformación en función de las diferencias climáticas. Una vez establecidos los coeficientes para la predicción de la abundancia de cada uno de los compuestos en función del clima, los cambios en los patrones pirolíticos de la SOM también se ilustraron a partir de una simulación de la composición molecular de la SOM para condiciones extremas de aridez o humedad.

Todos los resultados obtenidos confirman una clara relación entre la composición de la SOM y la capacidad de almacenamiento de C en los correspondientes suelos, e incluso puede reconocerse como dicha composición refleja el efecto de factores formadores del suelo como son la vegetación, clima o sustrato geológico. Del mismo modo, los constituyentes de la SOM definen una firma molecular ligada al avance de la desertificación. En particular, se ha demostrado que la diversidad de los metoxifenoles provenientes de la lignina parece aportar información acerca de la estabilidad de la SOM. Este también es el caso de los patrones de distribución de las series homólogas de otros compuestos relativamente resistentes a la degradación, como son los alcanos, que actúan a modo de registro molecular de cambios recientes en el funcionamiento del ciclo biogeoquímico de los suelos. Estudiando fracciones orgánicas más transformadas, como son los HAs, se comprueba la presencia de una firma molecular característica en cada suelo, que permanece en el tiempo y que se relaciona con el contenido de SOC y con los factores que han determinado su estabilización. Se ha observado que las estructuras más aromáticas u oxidadas se acumulan preferentemente en los suelos con escaso contenido en C, donde la SOM presenta una composición más heterogénea y compleja, comparativamente resiliente frente a futuros cambios ambientales. Por el contrario, la preservación de estructuras heredadas de la biomasa vegetal, como es el

caso de las ligninas alteradas, predomina en suelos con altos niveles de carbono.

SUMMARY

The progress of desertification, which is very pronounced in the Mediterranean area, and the emission of greenhouse gases into the atmosphere, especially CO_2, are taking the attention of the researchers. The establishment of the factors involved in these processes as well as the development of emergent technologies to solve these issues is a main objective in different scientific fields. In this perspective, soil conservation plays an important role, due to the content of soil organic matter (SOM) and its stability are important factors. Biogeochemical processes involved in the stabilization of soil organic carbon (SOC) are being the subject of study in this field. Some studies focus on the organo-mineral interactions, while others are interested in the relationship between the molecular structure or the SOM and its resilience. This thesis deals with the molecular characterization of SOM accumulated in different types of soil to establish its relationship with the potential for carbon storage in the corresponding soils, as well as the factors with a bearing on SOM quality. For this purpose, 35 soils with high variability in their organic carbon content (17–157 $g \cdot kg^{-1}$) have been selected. A detailed characterization of the organic matter has been carried out using destructive and non-destructive techniques such as analytical pyrolysis (Py-GC/MS), ^{13}C solid-state nuclear magnetic resonance (NMR), infrared (IR) spectroscopy, visible spectroscopy and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). In particular, the attention has been focused on certain types of biomarker compounds that could act as environmental indicators of soil biogeochemical processes. The families of alkanes and methoxyphenols were analyzed in detail; its molecular composition was used to distinguish between microbial synthesis and transformation of plant biomass.

This study examines the utility of the Shannon diversity index (*H'*), calculated from the abundance of alkanes analyzed by analytical pyrolysis (C_9–C_{31}) from whole soil samples, for the evaluation of the C storage potential and quality of SOM. A series of multivariate data treatments showed a significant relationship between the *H'* diversity of alkanes and the concentration of SOC. In particular, a significant relationship was found between SOC levels and the percentage of long-chain alkanes, while the percentage of short-chain alkanes was correlated with specific descriptors of SOM quality. Finally, partial least

squares regression (PLS) successfully predicted SOC content using exclusively the information provided by the alkane patterns.

In a parallel study, the molecular assemblages of methoxyphenols released by analytical pyrolysis from whole soil samples were also examined using the Shannon diversity index to describe the complexity of their pyrolytic patterns. A series of exploratory statistical methods (linear regression, PLS, multidimensional scaling (MDS), etc.) were applied to analyse the relationships between pyrolysis products and the chemical and spectroscopic characteristics of SOM and with the total SOC content. These results showed significant correlations between the progressive molecular diversity of the pyrolytic methoxyphenols and the SOC levels in the corresponding soils. The fact that the diversity of the phenolic signature provides information about the potential for carbon storage in soils can be interpreted as the progressive structural complexity of plant macromolecules modified by soil microorganisms, which makes them more difficult to be recognized by enzymes. From a quantitative point of view, PLS regression models based exclusively on the total abundance of the 12 major methoxyphenols were especially effective in predicting carbon storage in the soil.

After studying the information provided by analytical pyrolysis of SOM from different scenarios of carbon storage activity, the study was focused to the fraction traditionally considered most representative of the SOM, humic acids (HA). This fraction corresponds to a colloidal product of advanced transformation of plant and microbial biomass. The HA characterization was carried out using visible, IR and NMR spectroscopies. A PLS study using the intensities of digital IR spectra points (4000–400 cm^{-1}) as descriptors showed that there is a relationship between IR spectral patterns and SOC content. This was also the case with E4 index (i.e., indicative of progressive humification, and based on the optical density of HAs at 465 nm). The use of principal component analysis (PCA) and MDS suggested that the bands assigned to carbonyl and amide groups were characteristic in HA of soils with low C content, while the spectra of HAs from soils with high levels of C showed a typical pattern of lignin bands, which indicates accumulation of less transformed plant residues. The IR spectral patterns were analyzed in detail by digital treatments including weighted subtraction of spectra obtained by

averaging those of HAs from soils classified in the upper and lower quartiles according the SOC distribution, respectively, and calculating the statistical significance level of the differences. The results showed significant differences between the molecular composition of the HAs, according the SOC and E4 values. Peaks corresponding to aromatic, carboxyl and amide groups showed comparatively high intensity in HAs from soils with low SOC content, while peaks corresponding to lignin-derived structures were more marked in the spectra of the HAs from soils with high SOC content. In the second level of study of the HAs, its characterization was carried out using FTICR-MS. The application of PLS to FTICR-MS data showed a great potential for identifying the molecular components of HA that varies in terms of the SOC levels in different environments. A significant model to predict the SOC was obtained through PLS using as descriptors the 131 compounds detected by FTICR-MS in all HAs. In order to identify the compounds with the highest value as indicators of the SOC levels, the importance of the variables for prediction (VIP) was represented in the space defined by their atomic ratios using van Krevelen diagrams. The results showed a significant relationship between the molecular composition of HA and the C content stored in the soil: HA in soils with high levels of organic C displayed significantly higher proportions ($P < 0.1$) of molecular formulas corresponding to unsaturated lipids and lignin-derived compounds. On the other hand, low SOC levels were associated with comparatively higher proportions of saturated aliphatic structures.

The influence of the environmental factors involved in the total content and quality of SOM is poorly known, although it is closely related to the biodegradability of the SOM and the release rates of CO_2 into the atmosphere, which result in the progressive desertification of the soils. The objective is to study the molecular composition of SOM as a source of biogeochemical information on the impact of these conditions on the SOM characteristics in addition to the SOC levels. The 193 compounds released by pyrolysis from whole soil samples were included in a statistical study based on discriminant analysis to assess the impact of soil formation factors (climate, vegetation and geological substrate) on the total content and composition of the SOM. Van Krevelen diagrams were used for a perceptual recognition of characteristic patterns in the composition of SOM according to the soil formation factors, which showed a decreasing influence in the order: climate > vegetation >

geological substrate. The fact that the molecular composition of the SOM varied systematically according to environmental factors and the significant differences in SOM levels depending on the environmental scenarios suggests a qualitative and quantitative control of C sequestration in the soil.

The progressive depletion of SOM in Mediterranean ecosystems is associated to severe risks of erosion and desertification. Therefore, a study was conducted to establish the effect of climatic constraints on the composition of the SOM. The molecular composition studied by Py-GC/MS applied to whole soil samples was examined for its differences among soils developed in sites varying in the classical Emberger's pluviothermic quotient (*Q*). The application of PLS using the pyrolysis compounds as descriptors, allowed to predict ($P < 0.05$) the *Q* index, and subsequently to identify molecular proxies responsive to climatic variability. To illustrate the differences, the pyrolytic molecular assemblages from soils developed under extreme climatic conditions were compared using van Krevelen diagrams. The results showed that the molecular structure of the SOM retains environmental information related to the *Q* index, mainly reflected in the total abundance of methoxyphenols and alkylbenzene compounds. This fact suggests that the evolution of SOM is stabilized in more or less advanced stages of transformation depending on climatic characteristics. Finally, the differences between the proportions of the individual pyrolysis compounds in terms of the different bioclimatic scenarios were also determined from a simulation of the molecular composition of the SOM for extreme conditions of aridity or humidity.

The above results coincided in pointing out that a relationship exists between the composition of the SOM and the potential for C storage in the corresponding soils. It was also possible to describe how this composition reflects the effect of soil-forming factors such as vegetation, climate or geological substrate. In the same way, the SOM constituents define a molecular signature linked to the progress of desertification. For instance, the diversity of methoxyphenols from lignin is a valuable surrogate of carbon stability. This is also the case with the distribution patterns of the homologous series of compounds relatively resistant to degradation, such as alkanes, which act as a molecular record of recent changes in soil biogeochemical cycle. The above results are reinforced with those from the analysis of the

most transformed fractions of SOM, i.e., the HA with a characteristic molecular signature that remains in time and can be related to SOC levels and the factors involved in its stabilization. In particular, aromatic or oxidized structures accumulate preferentially in soils with low C content, whereas SOM has a more heterogeneous and complex composition in C-depleted soils, which could point to an enhanced resilience to future environmental changes. On the contrary, preservation of structures inherited from plant biomass, as altered lignins, is more characteristic in soils with high carbon levels.

Capítulo 1:
Introducción y Objetivos

1.1 El cambio global, la desertificación y el papel del suelo

En la actualidad el término "*cambio global*" o "*cambio climático*" está siendo frecuentemente contemplado en diversos ámbitos de la investigación (Lal, 2004; Janzen, 2006). A finales de este siglo se estima que la temperatura global del planeta se incremente entre 2–7 °C, por lo que se pronostica que la cantidad y distribución de las precipitaciones se verá alterada (Wu et al., 2011). La principal causa de todo ello se atribuye al gran incremento de la concentración de gases de efecto invernadero en la atmósfera, especialmente al dióxido de carbono (CO_2), de ahí el gran interés en reducir estas emisiones e intentar paliar este problema. La progresiva búsqueda de energías alternativas como las renovables y el desarrollo de tecnologías más eficientes para disminuir estas emisiones son bien conocidas, pero ¿es esto suficiente?. Junto a esta problemática del nivel de CO_2 en la atmosfera, no debemos olvidar uno de los factores íntimamente ligados como es la desertificación. Los procesos de degradación y pérdida de suelo en los ecosistemas mediterráneos han adquirido especial importancia en las últimas décadas. Esta desertificación promueve la destrucción de la estructura del suelo, la pérdida y alteración de su materia orgánica y por consiguiente la disminución de su fertilidad.

Uno de los mayores reservorios de carbono de la Tierra se encuentra en el suelo. El contenido de carbono orgánico en los suelos naturales del planeta se estima en un valor aproximado de 1.500 Pg en el primer metro de profundidad (Batjes, 1996), de los cuales aproximadamente el 50% de este carbono orgánico se encuentra en los primeros 30 cm de profundidad. A pesar de que los suelos sean uno de los mayores almacenes de carbono orgánico, el ecosistema terrestre es vulnerable a las variaciones ambientales a corto plazo. Un cambio en estos ecosistemas debido a la intervención de la actividad humana (incendios, deforestación, cambio de uso del suelo, etc.) podría conllevar cambios en el contenido y composición de la materia orgánica contenida en los suelos y la correspondiente degradación de los mismos (Smith, 2008; Jiménez-González et al., 2016; Pizzeghello et al., 2017). Estas ideas han llevado a que la comunidad científica preste especial atención a la cuantificación de estas cantidades de carbono y a las posibles

vías de estabilización del mismo. La conservación de los ecosistemas y el estudio de su evolución para entender la estabilización de carbono empiezan a cobrar más interés en diferentes campos de la comunidad científica, desde el estudio de la influencia de los componentes minerales del suelo (Kögel-Knabner et al., 2008; Manning et al., 2013; Wenting et al., 2014) hasta el del efecto de los microorganismos (Barré et al., 2018; Muñoz-Rojas et al., 2018).

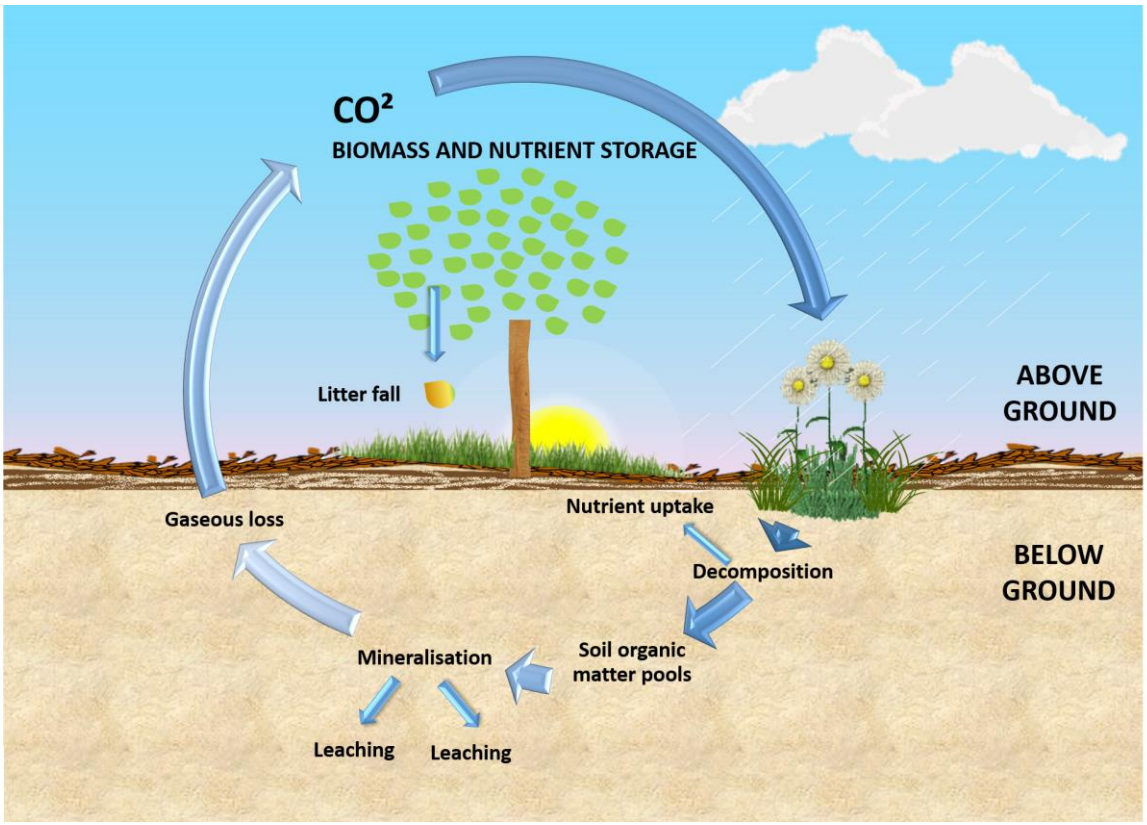


Figura 1.1. Esquema simplificado del ciclo del carbono.

Para entender mejor este fenómeno es necesario explicar el concepto del ciclo de carbono. En un modelo esquemático (Figura 1.1) se pueden observar las distintas etapas de transformación de los compuestos a los que se incorpora el carbono terrestre. El contenido final de este carbono almacenado en el suelo es el resultado de un balance entre la tasa de adición de carbono orgánico proveniente principalmente de la biomasa vegetal y la tasa de mineralización del mismo y por lo tanto su retorno a la atmósfera en forma de CO_2. En cuanto a este ciclo, la problemática radica en la medida en que se produce un incremento descontrolado o un desequilibrio en alguna de las etapas del mismo. En esta última parte es donde entra el factor humano, el

incremento de la actividad antrópica ha generado un aumento de la concentración de CO_2 en la atmósfera y una destrucción progresiva de los suelos alterando así este balance natural, generándose un aumento de la temperatura del planeta en las últimas décadas.

1.2 El secuestro de carbono en el suelo

La mayor parte de las investigaciones se han centrado en estudiar como los distintos suelos mantienen relativamente constantes sus niveles de carbono mediante diferentes interacciones que determinan las distintas vías de "*secuestro de carbono*" en el suelo, es decir pretenden identificar las distintas rutas de estabilización del carbono, así como los distintos factores implicados en esta estabilización.

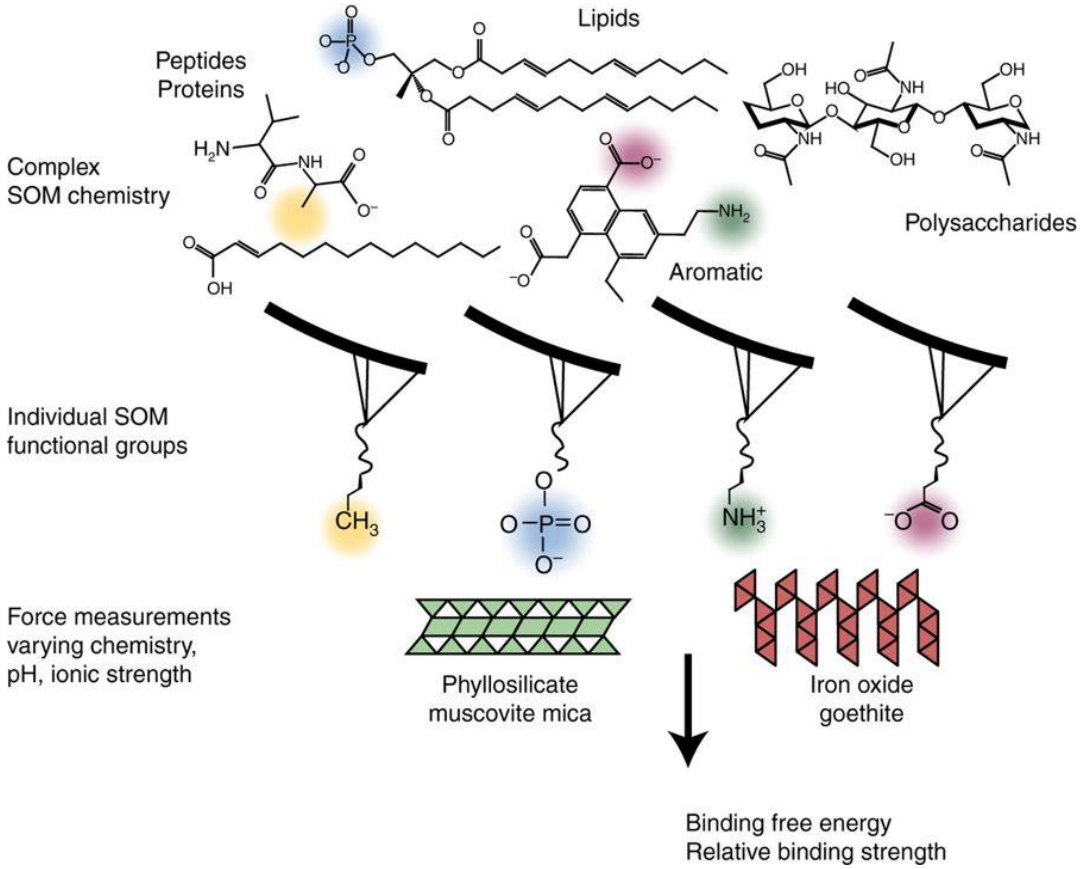


Figura 1.2. Modelo esquemático propuesto por Newcomb et al. (2017). Interacción entre grupos funcionales químicos que se encuentran normalmente en la materia orgánica del suelo (SOM) y dos minerales modelo: la moscovita y la goethita.

Muchas investigaciones focalizan su interés en la mineralogía del suelo (Kögel-Knabner et al., 2008; Wenting et al., 2014) mostrando como las interacciones órgano-minerales entre la materia orgánica del suelo (SOM) y la fracción mineral pueden estabilizar esta SOM, protegerla de la degradación o dejarla inaccesible para otras reacciones. En la Figura 1.2 se muestra un modelo de interacción órgano-mineral propuesto por Newcomb et al. (2017). Otros estudios le atribuyen mayor importancia a la protección física de la materia orgánica por los agregados del suelo, a través de la encapsulación de la misma por éstos. A través de un fraccionamiento físico del suelo, se realiza un estudio de las distintas fracciones de agregados (Dungait et al., 2011; Simonetti et al. 2017), observando como se acumulan las distintas formas de SOM dependiendo del tamaño de partícula de los agregados. Estos estudios sugieren una protección física de la materia orgánica en el interior de los agregados (Rabot et al., 2018), evitando que sea accesible a los microorganismos o protegiéndola de las posibles reacciones que pudieran degradarla en el suelo. Otros estudios ponen de manifiesto como el manejo del suelo puede condicionar en gran medida su contenido de carbono. Diversos trabajos (Pizzeghello et al., 2017; Hernández et al., 2018) se centraron en ver como el cambio de uso o manejo del suelo, incluyendo las prácticas de cultivo, podían influenciar la calidad y el contenido total de carbono. En efecto, los cambios experimentados en la estructura de la SOM por acción de diferentes tipos de perturbaciones ambientales son todavía poco conocidos, debido a la complejidad de su aislamiento y caracterización.

En particular, los procesos de estabilización de la materia orgánica asociados con cambios en su naturaleza química se encuentran aún escasamente estudiados. Con el presente trabajo se pretende contribuir al conocimiento de los factores que determinan la composición y estructura de la materia orgánica del suelo, así como su variabilidad en distintos escenarios ambientales. Ello implica intentar comprender los cambios estructurales, incluso a nivel molecular, experimentados por la SOM en el transcurso de su maduración, desde los primeros estados de transformación de la biomasa hasta las etapas más avanzadas de formación de sustancias húmicas, que condicionan su resistencia a la biodegradación, y los niveles resultantes de carbono en los suelos.

1.3 La materia orgánica del suelo

El suelo constituye un sistema muy complejo y dinámico en el que interactúan diversos constituyentes orgánicos y minerales dando lugar a su morfología y sus propiedades funcionales. De todos estos constituyentes, la materia orgánica presenta un especial interés, ya que a pesar de que algunas veces representa un porcentaje muy pequeño del total de suelo (< 1% en suelos semiáridos), ejerce una importante acción en sus propiedades, que se traducen en la calidad del suelo y la productividad de los ecosistemas (Lobo et al., 1984). Algunas de estas propiedades son bien conocidas y de enorme importancia como puede ser la estabilidad de la microestructura, en relación con la porosidad del suelo (Chen et al., 2017), la capacidad de intercambio catiónico (Brady, 1990), influencia en el pH del suelo (Brady, 1990), la capacidad de retención hídrica (Nath, 2014) o la repelencia al agua (Jordán et al., 2014; Jiménez-Morillo et al., 2016), sin olvidar la formación de agregados y la estabilidad de los mismos (Almendros, 2004; Shahbaz et al., 2017), que están íntimamente ligados a la materia orgánica. Debido a que todas las propiedades del suelo son afectadas en mayor o menor medida por los cambios cualitativos y cuantitativos experimentados por la SOM, pequeñas variaciones en su naturaleza pueden tener grandes repercusiones en la evolución del suelo en un tiempo relativamente corto. De hecho, la estructura de la SOM representa una huella de la influencia de la biota y factores físicos que han afectado la evolución de todo el ecosistema (Stevenson, 1994; Jiménez-González et al., 2017, 2018, 2019), de tal forma que cambios en la composición de la SOM se ven reflejados en la estabilidad del suelo.

El principal origen de la SOM es el material vegetal fresco que llega al suelo desde las plantas. Este material fresco junto con la biomasa microbiana, experimenta diversas transformaciones en su estructura química a lo largo del tiempo. Estas transformaciones dependen de diversos factores como pueden ser la climatología, la microfauna que se alimenta de este material fresco y favorece su transformación, y finalmente los microorganismos del suelo, que presentan una gran diversidad y actividad a través de procesos enzimáticos que permiten la transformación de este material vegetal. A todo este proceso de transformación de la materia orgánica, originalmente constituida por componentes de la biomasa hasta su evolución hacia formas transformadas

(con coloración oscura, elevada reactividad química y alta estabilidad frente a la degradación biológica) se le conoce como proceso de humificación.

La SOM se encuentra constituida por diferentes fracciones. Una de las menos evolucionadas y por lo tanto más parecida en su composición y estructura al material de partida es la denominada *materia orgánica libre*. Estaría compuesta principalmente por pequeños fragmentos de plantas. También presenta una gran importancia la fracción de *lípidos libres,* ya sean sintetizados por microorganismos o heredados de las ceras epicuticulares de las plantas, a pesar de encontrarse en baja proporción respecto a las demás fracciones. Es también el caso de los compuestos solubles procedentes de exudados de plantas (foliares o radiculares), o del metabolismo de los microorganismos, que ejercen importantes funciones en el suelo y son responsables de varias de sus propiedades. Otras fracciones de SOM que se encuentran en fases más avanzadas de evolución serían las denominadas *sustancias húmicas*, de naturaleza macromolecular extremadamente compleja debido a su composición heterogénea, por lo que su estructura continúa siendo en la actualidad objeto de diferentes investigaciones (Almendros, 2004; Hernández et al., 2019).

1.3.1 Las sustancias húmicas

Al ser la fracción más transformada de la materia orgánica del suelo, su estructura es muy heterogénea, lo que hace que su estudio sea particularmente complejo y requiera la aplicación de diversas técnicas analíticas que aportan información complementaria. Estas sustancias húmicas se pueden diferenciar en tres grandes grupos: *ácidos fúlvicos* (FA), *ácidos húmicos* (HA) y *humina*. Los ácidos fúlvicos serían la fracción de menor complejidad y tamaño molecular. Estos presentan elevada solubilidad cuando se encuentran libres, no precipitados en forma de fulvatos. Son solubles tanto en medios ácidos como básicos, por esta razón en los lixiviados de los suelos ácidos pueden presentar un color amarillento debido a su presencia en solución. La fracción humina es la fracción más difícil de aislar de las tres. Se trata de una fracción fuertemente asociada a los minerales de tipo arcilla y a los óxidos de hierro y aluminio, frecuentemente protegida en los agregados del suelo, lo que hace muy difícil tanto su aislamiento como su caracterización. Finalmente encontramos los ácidos húmicos, que constituyen

la fracción de la materia orgánica que solo es soluble en medio alcalino, pero se vuelve insoluble en agua a valores ácidos de pH o en presencia de cationes polivalentes. Su estructura nativa en el suelo aún es objeto de controversias, pues si bien mediante la extracción en medio alcalino puede obtenerse una preparación válida para conocer su composición, el material aislado no conserva su estructura tridimensional de cuando se encontraba enlazado con los constituyentes reactivos de la matriz del suelo. Por ello, se han postulado distintas teorías (Piccolo, 2002) que plantean su presencia en el suelo en forma de agregados supramoleculares resultantes de la agregación de micelas de tamaño relativamente reducido.

Estas sustancias húmicas también presentan una gran capacidad de formar complejos con los metales (Stevenson, 1994; Chechevatov et al., 2017). La fracción de ácidos húmicos presenta alto grado de transformación, en la medida que no se reconocen en ella grandes similitudes con las macromoléculas vegetales o microbianas aisladas. Precisamente por esta elevada variabilidad estructural, dependiente del material de partida y las condiciones ambientales, se considera que su composición podría estar ligada a todos los procesos y agentes involucrados en el ciclo del C del suelo, aportando valiosa información sobre la estructura y funcionamiento del ecosistema como podría ser su potencial de secuestro de carbono (Piccolo et al., 2018). A través del estudio de su composición se podrían identificar los tipos de suelos con materia orgánica más estable o recalcitrante, y por lo tanto de mayor capacidad de secuestro de carbono. Gran parte de esta Tesis doctoral se centrará en el estudio de esta fracción con la intención de establecer su posible relación con el contenido de carbono del suelo, como se desarrollará en los capítulos 5 y 6.

1.4 La lignina

Al hablar de SOM debemos tener en cuenta los precursores a partir de los cuales se forma. La vegetación es la principal fuente de materia orgánica; la biomasa vegetal está constituida principalmente por carbohidrato y lignina. Este hecho hace que la lignina sea uno de los grandes temas de investigación relacionados con la SOM (Tinoco et al., 2002; Jiménez-González et al., 2017).

1.4.1 Estructura de la lignina

La lignina es el constituyente más abundante en la vegetación terrestre después de los polisacáridos, encontrándose presente en todas las plantas leñosas. Esta macromolécula se encuentra en las paredes celulares de las células vegetales a las que aporta rigidez además de protección frente la desecación. Se trata de un polímero aromático de alto peso molecular y estructura heterogénea debido a la variabilidad en sus constituyentes estructurales, principalmente grupos fenólicos sustituidos con grupos metoxilo, y unidos entre sí por distintos tipos de enlaces, como se muestra en el modelo estructural propuesto por Stewart et al. (2009) representado en la Figura 1.3. Debido a esta heterogeneidad, la lignina es una macromolécula difícil de degradar, siendo los hongos los organismos más activos en sus etapas iniciales de degradación.

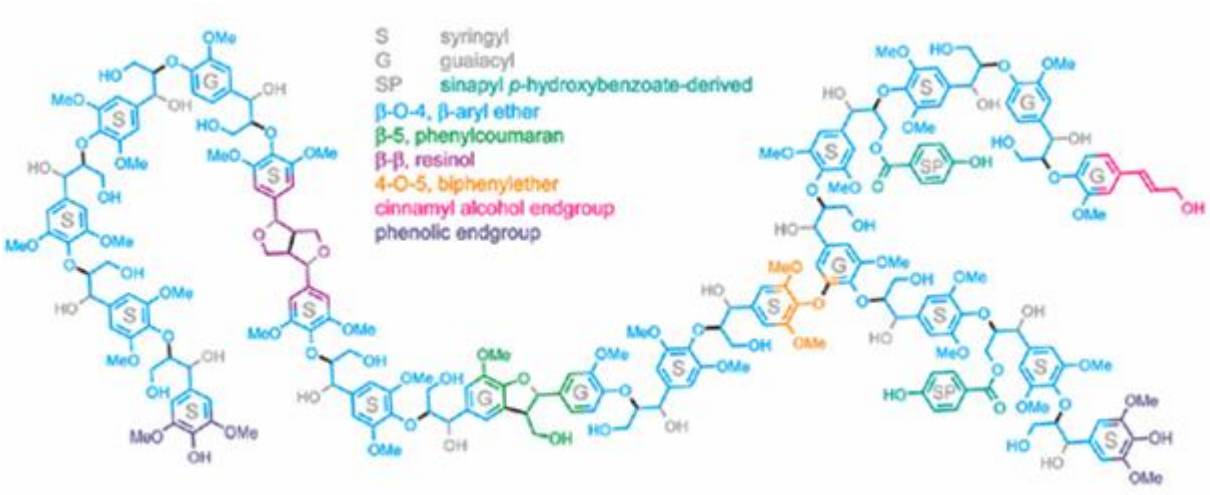


Figura 1.3. Modelo esquemático de la estructura de la lignina propuesto por Stewart et al., (2009) y Vanholme et al., (2010).

Las ligninas se forman esencialmente por la polimerización de tres tipos de unidades estructurales denominados monolignoles que serían los ácidos y alcoholes fenilpropílicos (cumarílico, coniferílico y sinapílico). Por condensación de estas unidades se forma un heteropolímero (Ralph 1999, Boerjan et al. 2003) con unidades que no se repiten mediante un determinado patrón regular, de ahí la alta complejidad en su estructura. De cualquier forma, los distintos tipos de lignina suelen definirse por las abundancias relativas de sus unidades características, de tipo *p*-hidroxifenilo, guayacilo y siringilo, que

se presentan en diferentes proporciones dependiendo del tipo de vegetación, como se muestra en la Figura 1.4.

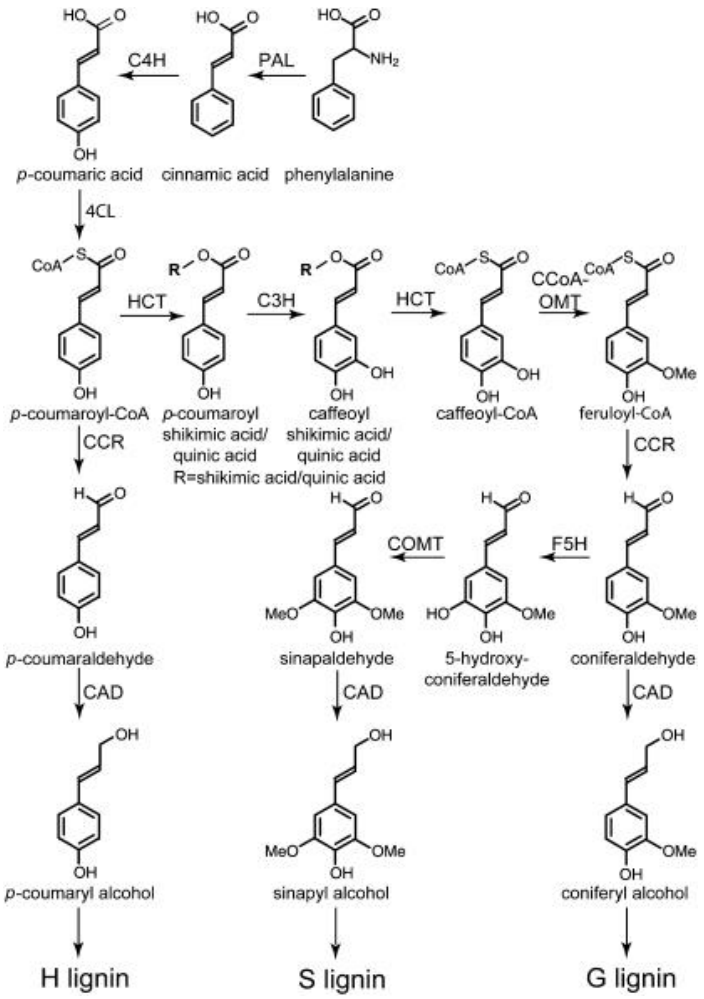


Figura 1.4. Esquema de reacciones de monolignoles que dan lugar a la macromolécula de lignina (Ralph 1999, Boerjan et al. 2003). Suelen ser reacciones de crecimiento en la que se acopla una nueva unidad de monolignol (cumarílico, coniferílico y sinapílico). PAL, Fenilalanina amonio liasa; C4H, cinamato 4-hidroxilasa; 4CL, 4-cumarato:CoA ligasa; C3H, *p*-cumarato 3-hidroxilasa; HCT, hidroxicinamoil shikimato/quinato transferasa; CCoAOMT, cafeoyl-CoA *O*-metiltransferasa; CCR, cinamoil-CoA reductasa; F5H, ferulato 5-hidroxilasa; COMT, ácido caféico *O*-metiltransferasa; CAD, cinamil alcohol deshidrogenasa.

Como se verá en el siguiente apartado, los patrones definidos por los distintos tipos de unidades estructurales de la lignina, especialmente los de tipo guayacilo y siringilo, son de gran utilidad para identificar el tipo de lignina y constituyen un marcador de la misma. El estudio de la lignina y su composición se ha llevado a cabo por medio de numerosas técnicas analíticas como pueden ser la espectroscopia infrarroja, resonancia magnética nuclear o

la degradación química o pirolisis seguida por cromatografía de gases, que serán empleadas en este estudio.

1.4.2 La lignina y la materia orgánica del suelo

Una vez conocida la estructura básica de esta macromolécula natural presente en las plantas vasculares, se debe tener en cuenta que su composición y estructura comienza a modificarse en el suelo mediante diferentes procesos que dependen de los factores ambientales. Durante esta transformación, la materia orgánica fresca empieza a perder progresivamente su estructura biosintética original, de tal forma que su seguimiento analítico constituye una herramienta de gran utilidad para evaluar su grado de transformación, fundamentalmente reconocible en los cambios en los patrones de las unidades de tipo metoxifenol. Estos compuestos son fáciles de identificar por diferentes técnicas analíticas, por lo que pueden ser utilizados para tener una idea de las etapas más o menos avanzadas de la humificación de la lignina en un determinado suelo. Una de las técnicas más utilizadas para estudiar la materia orgánica es la pirólisis analítica (Martín et al., 1979; De la Rosa et al., 2008a, 2012; González-Pérez et al., 2008; Miralles et al., 2015; Jiménez-González et al., 2016). Mediante esta técnica es posible identificar los 12 principales metoxifenoles, incluso analizando directamente muestras de suelo, sin necesidad de pretratamientos que pueden alterar o fraccionar la muestra. Los principales metoxifenoles liberados e identificados por pirólisis analítica son los derivados: 4-*H*-, metil-, etil-, vinil-, propenil- y aceto- de las estructuras de tipo guayacilo y siringilo (Figura 1.5). Además, diversos estudios han mostrado que diferentes tipos de plantas presentan una composición muy variable de estos metoxifenoles (Fengel & Wegener, 1984; Tinoco et al., 2002; Jiménez-González et al., 2016). Las plantas angiospermas presentan ligninas formadas por ambos tipos de metoxifenoles (tipo guayacilo y siringilo), mientras que las gimnospermas presentan mayor proporción de unidades de tipo guayacilo (Jiménez-González et al., 2017). Por todo ello, la variabilidad de estos compuestos en la materia orgánica de los suelos podría aportar importante información acerca de los agentes activos en los procesos de secuestro de carbono, como se estudiará en el capítulo 4.

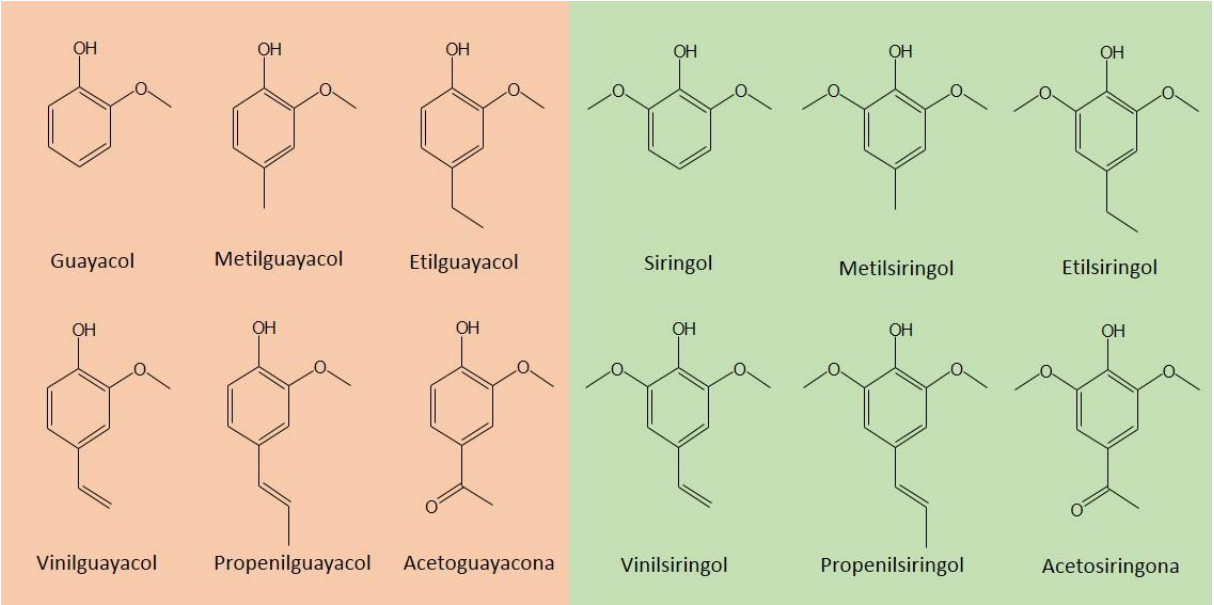


Figura 1.5. Estructura de los 12 metoxifenoles característicos de la estructura de la lignina identificados a partir de la materia orgánica del suelo por pirolisis acoplada a cromatografía de gases.

No solo la cromatografía permite evaluar la variabilidad en la composición de las formas alteradas de lignina en los suelos. Técnicas espectroscópicas como la resonancia magnética nuclear (NMR) o la espectroscopia infrarroja (IR) aportan información acerca de su presencia y grado de transformación en la SOM. En el caso de la espectroscopia IR se han realizado numerosos estudios sobre la lignina, estableciéndose las bandas diagnósticas de la misma (Farmer & Morrison, 1960; MacCarthy & Rice, 1985; Yang et al., 2011; Fernández-Getino et al., 2013; Miralles et al., 2015; Jiménez-González et al., 2019). En particular se considera especialmente indicativo de la presencia de lignina poco alterada la presencia de un conjunto de bandas sobre 1620, 1510, 1460, 1420, 1270, 1230, 1130 y 1030 cm^{-1} (Farmer & Morrison, 1960; Fengel & Wegener, 1984; Yang et al., 2011). La Figura 1.6 muestra un ejemplo de espectro IR de un ácido húmico con las bandas características marcadas. Determinaciones cuantitativas de la intensidad de estas bandas pueden ser muy útiles para el estudio de la resiliencia y madurez del humus (MacCarthy & Rice, 1985; Fernández-Getino et al., 2013; Jiménez-González et al., 2019).

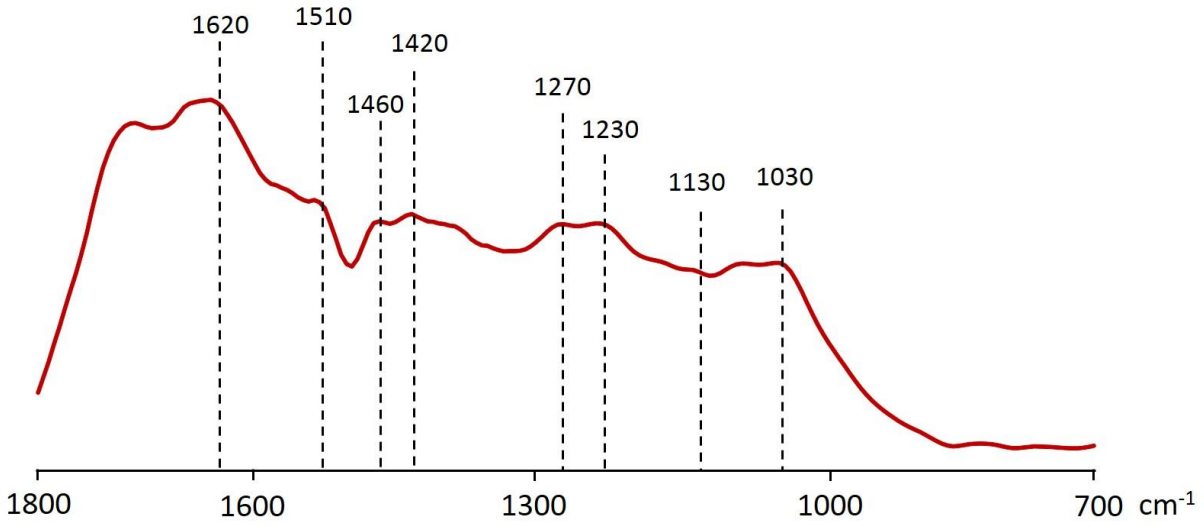


Figura 1.6. Ejemplo de espectro de infrarrojo (IR) de un ácido húmico (HA) extraído del suelo. Las bandas más representativas de la lignina (1620, 1510, 1460, 1420, 1270, 1230, 1130, 1030 cm^{-1}) se encuentran marcadas.

1.5 Los alcanos

Tanto en sus formas libres (fracción lipídica) como incorporados en la estructura macromolecular de las sustancias húmicas, los hidrocarburos saturados, o alcanos constituyen unos componentes constantes de la composición de la materia orgánica de todos los suelos. Estos compuestos, constituidos únicamente por C y H con una fórmula general de C_nH_{2n+2} se encuentran en multitud de organismos, desde las bacterias y hongos hasta las plantas y fauna del suelo (Montiel-Rozas et al., 2017; Tian et al., 2017). Los alcanos pueden ser una fuente de información sobre el origen y los procesos involucrados en la transformación de la SOM (Jansen and Wiesenberg, 2017); en cierto modo su distribución (abundancias relativas de los diferentes compuestos homólogos de las series que, generalmente entre el alcano C_9 y el C_{35} aparecen en todos los suelos con distribución muy diferente en función de su origen y grado de evolución) podría aportar valiosa información sobre el origen del carbono en el suelo y su estabilidad (Nip et al., 1986; Kögel-Knabner and Hatcher 1989; Almendros et al., 1991; Baldock et al., 1992). Con respecto a la humificación o transformación de la materia orgánica, diversos trabajos han mostrado como los alcanos de plantas o microorganismos acaban incorporándose a la estructura de las sustancias húmicas en el suelo

(Requema et al., 1996; Almendros & Dorado, 1999; Pendall & King, 2007; Jiménez-González et al., 2019).

La importancia de esta familia de compuestos presentes en la SOM se pone de manifiesto en relación con importantes propiedades físicas de los suelos. Una fracción considerable de alcanos se encuentra libre en la fracción lipídica del suelo y por su estructura presentan un gran carácter hidrofóbico, que condiciona gran parte de las propiedades físicas del suelo. Este carácter hidrofóbico hace que estos lípidos, impermeabilicen y favorezcan la estabilidad de los agregados, haciéndolos más resistentes (Jambu et al., 1983, 1985; Dinel et al., 1991). En casos extremos, el alto contenido en alcanos puede influir en la repelencia al agua de los suelos. Esta característica puede provocar que la infiltración del agua en el suelo sea muy baja o incluso que se produzcan flujos superficiales del agua de escorrentía que favorezcan la erosión. Este fenómeno ha sido relacionado con el contenido de lípidos en el suelo (Rumpel et al. 2004; Jordán et al. 2013; Jiménez-Morillo et al. 2016). Por último, los alcanos llegan a incorporarse a la materia orgánica más transformada, formando así parte de la estructura de los ácidos húmicos y acumulándose también como carbono orgánico estabilizado en el suelo (Amblès et al. 1991; Eglinton & Logan 1991; Nebbioso et al., 2015).

Debido a esta doble importancia de los alcanos—como fuente de información sobre procesos biogeoquímicos de los suelos y como agente activo en las propiedades físicas de los mismos—se ha decidido prestar especial atención a su composición, en la medida que pudiera informar directa o indirectamente sobre los procesos que determinan el almacenamiento de carbono en los suelos y la calidad o estabilidad de la materia orgánica correspondiente (Jiménez-González et al., 2018), como se verá en el capítulo 3.

1.6 Los factores formadores del suelo

La formación del suelo es un fenómeno que tiene lugar en prolongados periodos de tiempo, en tanto que la pérdida o degradación del mismo, puede tener lugar en un corto periodo, sobre todo cuando se debe a la incidencia de fenómenos climáticos catastróficos. Los factores que presentan un papel

importante en la formación de los suelos son muy diversos (Jenny, 1994; Towett et al., 2015). Estos factores incluirían cinco principales: clima, sustrato geológico, relieve, biota y el tiempo (Figura 1.7). El efecto de estos factores suele quedar reflejado en la composición de la materia orgánica (Jenny, 1941; Duchafour & Jacquin 1975) y en su contenido total (Johnson et al. 2011) como se verá en el capítulo 7.

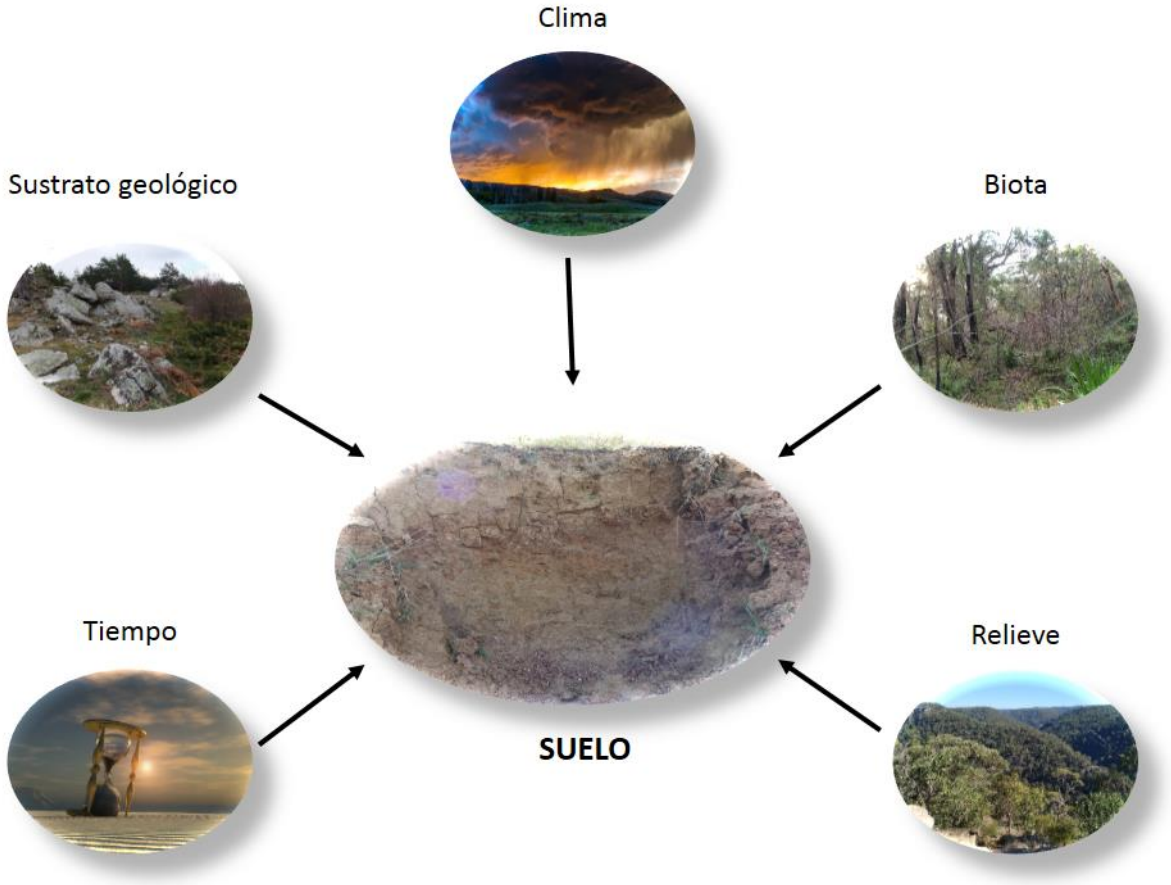


Figura 1.7. Los cinco principales factores formadores del suelo.

1.6.1 El clima

Históricamente se ha prestado especial atención a la climatología como uno de los principales factores que influyen en la formación de un suelo. Los factores climatológicos como las precipitaciones abundantes y la variabilidad en la temperatura a lo largo de las distintas estaciones, desempeña un papel importante en la edafogénesis. Es bien conocido que las condiciones de humedad favorecen la actividad microbiana en el suelo, que desempeña un

papel primordial en las transformaciones de la materia orgánica. Por otro lado, las temperaturas extremadamente bajas producen un efecto de aletargamiento o hibernación de los microorganismos del suelo, disminuyendo así su actividad y la velocidad de muchas reacciones químicas. Por supuesto, también se debe mencionar el efecto directo de la humedad y la temperatura en la meteorización del sustrato geológico, mediante procesos físicos a bajas temperaturas y mayor influencia de las reacciones químicas a temperaturas más elevadas.

1.6.2 El sustrato geológico

El material geológico o roca parental es otro de los factores que condicionan la formación de un suelo. Este sustrato geológico es la plataforma donde se desarrolla el suelo y podría ser considerado como un elemento pasivo que se ve afectado por los otros factores formadores. A pesar de existir un número limitado de tipos de rocas, estas pueden dar lugar a diversos tipos de suelos en combinación con los otros factores. Las rocas se pueden clasificar en tres grandes grupos: ígneas, sedimentarias y metamórficas. Las rocas ígneas son aquellas formadas a partir de mezclas de minerales fundidos (magma) y pueden ser intrusivas (plutónicas) cuando se forman lentamente a gran profundidad como es el caso del granito, o extrusivas (volcánicas) formadas por enfriamiento más rápido y en la superficie terrestre, por efecto de las erupciones volcánicas. Las rocas sedimentarias se originan mediante la disolución o transporte y posterior acumulación de fragmentos de otras rocas. Estos sedimentos se acumulan en las cuencas y a través de la diagénesis dan lugar a rocas sedimentarias consolidadas. Finalmente, las rocas metamórficas se forman a través de una transformación de los minerales de otras rocas. El proceso de transformación, debido al efecto de agentes como la presión o temperatura se conoce como metamorfismo. Algunos ejemplos de estas rocas son los gneises, pizarras....

1.6.3 La biota

El siguiente de los factores es la biota: que engloba los organismos que se desarrollan en el suelo e intervienen en su formación. La biota compuesta tanto por animales como por vegetales y microorganismos, contribuye tanto en la formación de material de partida, como en la fragmentación, y trasformación química de los componentes del suelo. La vegetación con sus

raíces penetra por las grietas de las rocas y las fragmenta. El aporte de material vegetal al suelo y su posterior transformación por otros organismos representa una fuente primaria de SOM. Son numerosos los estudios clásicos que se han centrado en establecer como distintos tipos de vegetación pueden influir en la calidad y cantidad de la SOM (Duchaufour & Jacquin, 1975; Lopez-Capel et al., 2008).

1.6.4 El relieve

El relieve del terreno es otro de los cinco factores involucrados en la formación del suelo. El relieve tiene una gran influencia en la erosión y en la sedimentación en función de la morfología del terreno. Dependiendo de la pendiente, los suelos pueden ser vulnerables a la erosión hídrica, o en cambio, en las llanuras o depresiones, los procesos de acumulación pueden verse favorecidos. Aparte de este efecto, la orientación de las laderas juega un papel muy importante; la orientación de la ladera puede implicar que el suelo se encuentre en zona de umbría o por el contrario expuesto a mayor irradiación de luz solar. Por lo tanto, influye de manera indirecta en la temperatura del suelo, y a su vez en el periodo en que permanece húmedo. Por todo ello, el relieve presenta un importante efecto sobre todos los procesos de formación el suelo.

1.6.5 El tiempo

El último factor, pero no por eso el menos importante, es el tiempo. La formación de un suelo es un proceso lento que necesita cientos o miles de años, bastante más tiempo que el que puede ser necesario para destruirlo. Debido a esto, el tiempo se considera como un factor de formación.

1.7 Objetivos

El objetivo general de esta Tesis Doctoral sería establecer los factores que determinan la capacidad de almacenamiento de C de los suelos: explicar porque unos suelos tienen mayor contenido de SOM que otros. Igualmente, en este trabajo, se pretende aportar información sobre la diferente calidad de la materia orgánica—establecida en las diferencias en su composición molecular—en función de su contenido y que permita valorar las ventajas e

inconvenientes ambientales del almacenamiento de carbono por parte de los suelos. Con este propósito, se han planteado diversos objetivos más específicos:

- Un primer objetivo sería estudiar la información biogeoquímica suministrada por la composición molecular de alcanos encontrados en la SOM. Al ser compuestos relativamente abundantes y estables a lo largo del tiempo, se espera que su caracterización proporcione datos acerca del origen y mecanismos de estabilización de la SOM (capítulo 3).
- La SOM tiene su principal origen en la vegetación que se desarrolla sobre el suelo donde sus restos son transformados en el transcurso de su biodegradación y humificación. Por este motivo se plantea estudiar la evolución de la lignina a través de la caracterización molecular de sus unidades estructurales. Para su identificación se aplicará pirólisis analítica, que libera una mezcla de metoxifenoles característica para cada suelo, que a su vez se intentará relacionar con la incidencia e intensidad de las vías de transformación de la SOM (capítulo 4).
- A lo largo de todo el estudio, se plantea comprobar si la composición molecular de los ácidos húmicos, una de las fracciones más estables y con composición diferencial y característica de los distintos suelos, presenta diferencias sistemáticas en función del contenido total de SOM. Ello permitiría, por otra parte, establecer objetivamente los niveles de calidad de humificación, entendida como la medida en que la composición molecular de la fracción de HA difiere con respecto a la de la biomasa originaria aportada al suelo (capítulos 5 y 6).
- Evaluar la medida en que los factores formadores del suelo (clima, vegetación y sustrato geológico) pueden influir en el contenido de SOM, y describir posteriormente cuales serían las características diferenciales en la composición de la SOM como consecuencia del impacto de cada uno de los factores estudiados (capítulo 7).
- Finalmente, explorar la medida en que la variabilidad de los principales factores climáticos (humedad, temperatura, déficit hídrico…) que condicionan el avance de los procesos de desertificación, se refleja en la composición molecular de la correspondiente SOM, al mismo tiempo

que influye en el diferente contenido de carbono orgánico almacenado en los suelos (capítulo 8).

Capítulo 2:
Materiales y Métodos

2.1 Área de estudio y muestreo

Se seleccionaron 35 puntos de muestreo de toda España (Figura 2.1) con la finalidad de comparar una gran variedad de suelos con amplio rango en el contenido en carbono orgánico. El muestreo se realizó tomando la capa superficial del suelo (*topsoil*) (0–10 cm), en la que se encuentra la mayor concentración de carbono. El muestreo se realizó retirando previamente la capa de hojarasca, en caso de encontrarse presente. Se recogieron tres muestras de suelo de unos 500 g en cada punto de muestreo y se realizó una combinación de estas tres muestras. Finalmente, se secaron al aire y se homogeneizaron pasándolas a través de un tamiz de 2 mm.

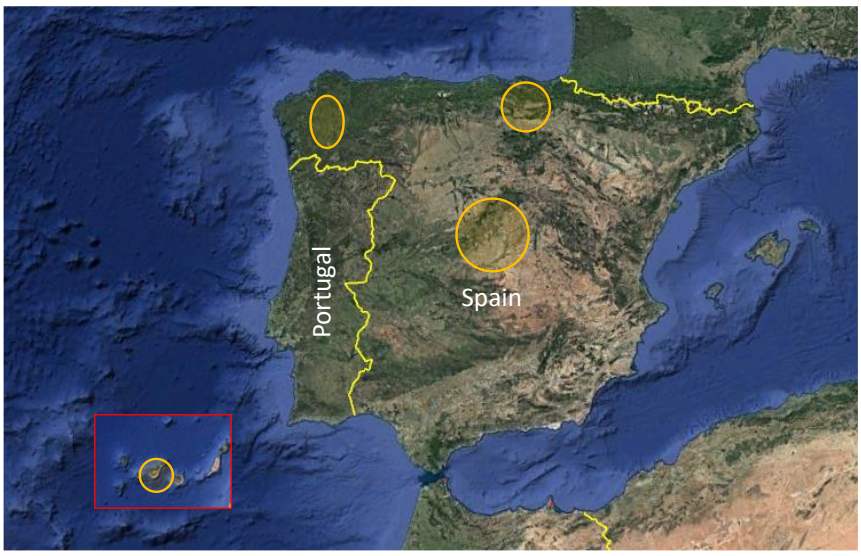


Figura 2.1. Localización de las áreas de muestreo marcadas con círculos amarillos. Imagen obtenida de Google Earth.

Los suelos fueron clasificados de acuerdo con la Organización de las Naciones Unidas para la Alimentación y la Agricultura (WRB 2014) y de acuerdo con el Departamento de Agricultura de Estados Unidos (USDA 2014). En función de la climatología, los suelos se encontraban bajo tres principales tipos climáticos de la clasificación de Köppen (Kottek et al., 2006): i) climatología de tipo Cfa, clima templado, precipitaciones todo el año y con una temperatura media del mes más cálido >22 °C; ii) Cfb, clima templado, precipitaciones todo el año pero con temperatura media del mes más cálido <22 °C >10 °C; iii) Csa, clima templado con estación seca en verano y con temperatura media del mes más cálido >22 °C. En las Tablas 1 y 2 se muestran las coordenadas de los puntos

de muestreo y las características generales de los correspondientes ecosistemas: vegetación predominante, sustrato geológico, climatología.

2.2 Análisis granulométrico

La textura de un suelo es una propiedad física muy importante. La distribución del tamaño de partículas, expresado como su contenido en arcilla, limo y arena puede condicionar en gran medida las demás propiedades físicas y fisicoquímicas del suelo, e influir en el contenido de otros componentes como puede ser el de los iones solubles o la materia orgánica del suelo. La textura se encuentra, a su vez, asociada a otras propiedades como son la estructura del suelo o la retención hídrica del mismo. En el presente estudio se ha utilizado el criterio de la FAO (Tabla 2.2) para la definición de los tamaños de partícula del suelo.

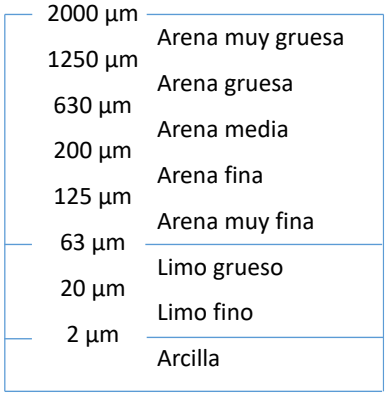


Figura 2.2. Tamaños de partícula establecidos por el sistema de clasificación de la FAO (IUSS Working Group WRB, 2014).

El análisis se llevó a cabo mediante el método densimétrico propuesto por Bouyoucos (Bouyoucos, 1936). Para ello, las muestras de suelo, previamente tamizadas por 2 mm (“tierra fina”) se sometieron a un tratamiento con H_2O_2 11% (p/v) (36 vol. O_2) a una temperatura de 60 °C, con objeto de eliminar la materia orgánica, que podría conducir a considerables errores en el análisis granulométrico. Una vez que las muestras estaban libres de materia orgánica se tomaron 50 g y se les añadió 400 cm^3 de agua y 20 cm^3 de disolución dispersante. La disolución dispersante fue preparada añadiendo 35.7 g de

$(NaPO_3)_6$ (hexametafosfato sódico) y 7.94 g de Na_2CO_3 enrasando a 1 L. La mezcla de la muestra con la disolución dispersante se agitó mecánicamente durante 15 min, luego se colocaron en probetas de 1 dm^3 y se les adicionó agua hasta completar 1 dm^3.

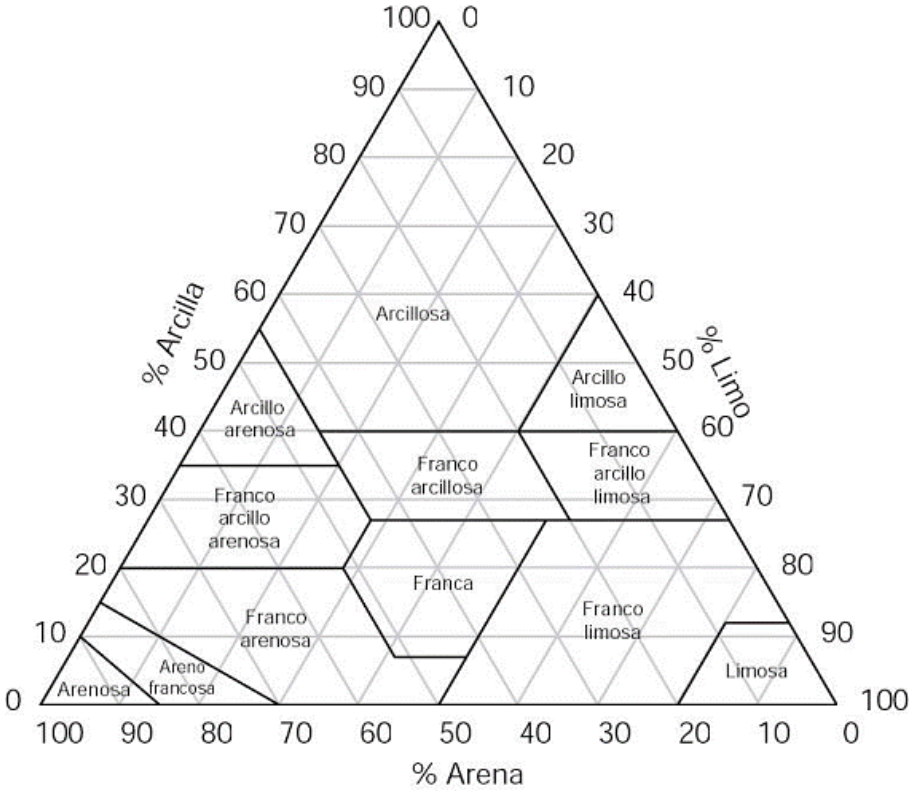


Figura 2.3. Triángulo de clasificación de texturas de suelo según la FAO (IUSS Working Group WRB, 2014).

Una vez que todas las suspensiones de suelo se encontraban en las probetas se agitaron manualmente con una varilla terminada en un disco perforado para mantener toda la muestra en suspensión. Se realizaron medidas de densidad y temperatura a los 40 s y a las 2 h después de la agitación manual, aplicándose factores de una tabla de conversión para obtener los valores de arcilla, limo y arena. Los resultados obtenidos correspondientes a las distintas fracciones fueron representados en un diagrama triangular (Fig. 2.3) con escala de valores para los distintos tipos texturales, obteniendo así el tipo textural de cada suelo (Tabla 3).

Figura 2.4. Determinación de la textura del suelo. Eliminación de la materia orgánica y posterior aplicación del método de Bouyoucos para determinar textura. Separación de las arenas mediante el empleo de tamices.

Para la determinación de los tamaños de arena, se recuperó la fracción arena del ensayo por Bouyoucos. Se colocó en un tamiz de 63 µm de luz y se tamizó con agua para eliminar las fracciones limo y arcilla. Posteriormente la arena se secó en estufa a 90 °C y a continuación se separó a través de varios tamices (1250, 630, 200, 125 y 63 µm) separando en cada uno las distintas fracciones de arena. El contenido de las fracciones, determinado gravimétricamente, fue expresado en $g \cdot kg^{-1}$ de suelo.

2.3 Determinación de la retención hídrica

La capacidad del suelo para retener agua influye en gran medida en el desarrollo de la vegetación y en la actividad biológica general. Esta capacidad del suelo se encuentra estrechamente asociada a su estructura, y sobre todo, a su contenido en materia orgánica, toda vez que la cantidad de materia orgánica y arcilla favorecen en gran medida la retención hídrica de un suelo. En función de las fuerzas que mantienen absorbida el agua en el suelo, pueden diferenciarse varios tipos:

- El *agua gravitacional* sería el agua que se pierde por percolación, no siendo retenida en ningún momento por el suelo y pasando rápidamente a las capas profundas.
- El *agua capilar absorbible*, queda retenida en el suelo después del periodo de lluvia y es absorbible por la planta.

- El *agua capilar no absorbible y agua higroscópica* sería el agua que queda retenida en los poros más pequeños de 0.2 µm del suelo, no siendo posible su absorción por la planta. Sería el agua que sigue reteniendo el suelo después de aplicar una presión de 1.52 MPa (15 atm). El agua *útil* sería aquella retenida en poros con un tamaño de entre 0.2 y 0.8 µm.

El potencial capilar (pF) se utiliza para medir la fuerza de retención del agua por el suelo. *El punto de marchitez o marchitamiento* es otro concepto importante: es aquel en el que las raíces de las plantas no pueden generar suficiente presión osmótica para absorber el agua del suelo, y se clasifica en dos tipos: i) *punto de marchitez reversible o temporal,* que se alcanza cuando se marchita la planta, pero puede recuperarse añadiendo agua, ii*) punto de marchitez irreversible o permanente* en el cual la planta marchita no puede recuperarse (pF de 4.2) y corresponde al contenido en agua que permanece en el suelo tras aplicar una presión de 1.52 MPa.

En las distintas muestras de suelo, pasadas por un tamiz de 2 mm, se llevaron a cabo las determinaciones de la capacidad de retención de agua a la presión de 0.1 MPa (1 atm) y en el punto de marchitez permanente a 1.52 MPa.

Para la determinación de la retención hídrica del suelo a la presión de una atmósfera (0.1 MPa) se emplearon unos cilindros con orificios en la base de 2 mm de diámetro. Estos cilindros fueron llenados con un peso conocido de muestra de suelo. Los cilindros se sumergieron en agua hasta alcanzar un nivel de 2 cm sobre su base con el fin de que el agua humectara el suelo por capilaridad. Las muestras permanecieron durante 16 h en el baño de agua para garantizar la humectación total. Luego se sacaron del baño y se pesaron directamente (capacidad máxima). Posteriormente se dejaron drenar los cilindros durante 2 h y se volvieron a pesar. Una vez pesadas las muestras se secaron en estufa a 60 °C durante 24 h y luego se volvieron a pesar, para obtener el peso de suelo seco original. Por diferencia entre el peso de suelo húmedo y seco se obtiene el peso de agua retenida por el suelo (Guitián y Carballas, 1976).

Figura 2.5. Determinación de retención hídrica del suelo a la presión de 0.1 MPa. Empleo de un extractor Richards de presión-membrana para la determinación del punto de marchitez (1.52 MPa).

El *punto de marchitez* se determinó mediante un extractor de presión-membrana (de Richards). Las muestras fueron dispuestas durante 16 h en agua para humectarlas por completo. Posteriormente se introdujeron en el extractor de Richards y se sometieron a una presión positiva de 1.52 MPa. La membrana porosa, permeable al agua, pero no al aire, permite la infiltración del contenido en agua desplazada por la presión. Una vez expulsada el agua, se despresuriza lentamente el extractor y las muestras son extraídas y pesadas rápidamente, para luego desecarse en estufa a 105 °C durante 24 h, tras lo cual se vuelven a pesar. La diferencia de peso obtenido entre las muestras húmedas después de ser sometidas a la presión de 1.52 MPa y las muestras secas corresponde al agua capilar retenida en el suelo, que es lo que se conoce como contenido hídrico en el *punto de marchitez*. El *agua útil* se calcula por diferencia entre el valor obtenido en el primer ensayo a 0.1 MPa y el valor obtenido en el *punto de marchitez* (1.52 MPa).

2.4 Determinación del carbono orgánico y del nitrógeno total

La determinación de carbono orgánico total del suelo se llevó a cabo mediante el método de oxidación en húmedo (Walkley & Black, 1934). Este procedimiento se basa en la oxidación en medio ácido empleando dicromato potásico ($K_2Cr_2O_7$) 1*N* como oxidante. Posteriormente el exceso de $K_2Cr_2O_7$ que no haya reaccionado se valora con sal de Mohr 0.5*M* ($Fe(NH_4)_2(SO_4)_2$) empleando difenilamina como indicador de la valoración. Para esta determinación se emplearon de 0,2 a 0,5 g de suelo pulverizado, los cuales fueron depositados en un matraz Erlenmeyer de 500 cm^3. Posteriormente se le añadió 10 cm^3 de $K_2Cr_2O_7$ 0,1*N* y 20 cm^3 de H_2SO_4 concentrado. Finalmente se agitó ligeramente durante 1 min y se dejó en reposo 30 min. Se añadieron 100 cm^3 de agua destilada y se dejó enfriar. Una vez fría, se añadieron 10 cm^3 de H_2PO_4 concentrado (15*M*) y 1 cm^3 de difenilamina y se procedió a la valoración del exceso de $K_2Cr_2O_7$ empleando sal de Mohr 0.5*M*.

La determinación del nitrógeno total se realizó por el método de Kjeldahl (Prince, 1945). En este método se emplea H_2SO_4 concentrado (96%) junto con un catalizador compuesto por 100 g de K_2SO_4 y 5 g de Se pulverizado. Estos reactivos reaccionan con el N orgánico del suelo generando $(NH_4)_2SO_4$ que es valorado posteriormente permitiendo conocer la cantidad de nitrógeno.

2.5 Determinación del pH

El pH es una propiedad muy importante de los suelos, principalmente por su relación con la movilidad de los iones y por consiguiente su biodisponibilidad. Es un factor dependiente de factores como pueden ser la composición mineralógica del suelo, y el contenido y propiedades de la materia orgánica, de tal forma que valores bajos de pH pueden ser debidos al elevado contenido en grupos funcionales ácidos de la SOM (carboxilo, OH-fenólicos o Al^{3+} en posiciones de cambio.

La determinación de pH se llevó a cabo en una suspensión de suelo (previamente tamizado a 2 mm) en agua, en una relación de 1:2,5 (w:w). La mezcla se agitó durante 1 h y posteriormente se dejó reposar 30 min. Los

valores de pH se determinaron con un pH-metro modelo *pH7* de *XS Instruments*. Resultados representados en al Tabla 4.

2.6 Determinación de la conductividad eléctrica

La conductividad eléctrica (EC) proporciona una idea del contenido en sales solubles presentes en un determinado suelo. Su determinación es muy útil para identificar posibles problemas de salinidad, que efecta al crecimiento de la vegetación o la germinación de las semillas. Los suelos muy salinos presentan problemas de absorción de agua para las plantas, así como frecuentes problemas de baja agregación y pobre estructura.

La medida de la conductividad eléctrica se realizó preparando una suspensión de suelo, previamente tamizado por 2 mm, en agua, en una proporción 1:5 (p:p). La mezcla se agitó durante 1 h y el líquido sobrenadante se separó por filtración. La conductividad se midió en la solución sobrenadante con un conductímetro modelo COND7 de XS Instruments. Los valores de EC obtenidos están indicados en la Tabla 4.

2.7 Determinación de la capacidad de intercambio catiónico y de los cationes de cambio

La capacidad de intercambio catiónico (CEC) informa acerca de la concentración de iones adsorbidos en el complejo de cambio de un determinado suelo, expresada como $cmol_c \cdot kg^{-1}$. Los cationes mayoritarios en los suelos se denominan *bases de cambio* y corresponden al Ca^{2+}, Mg^{2+}, Na^{+} y K^{+}. La suma de sus concentraciones se representa por S (suma de bases). El contenido en estos cationes y su biodisponibilidad desempeñan un papel muy importante en las propiedades de los suelos. Los cationes de cambio se encuentran retenidos en la matriz del suelo, conjunto de superficies reactivas del mismo, principalmente constituida por la materia orgánica humificada, oxihidróxidos y fracción arcilla, que poseen los grupos funcionales cargados negativamente y con capacidad de interaccionar con estos cationes.

La extracción de los cationes se llevó a cabo mediante un desplazamiento de los mismos por NH_4^{+} de forma que pasen a la solución y el amonio quede

retenido en el complejo de cambio del suelo (Rhoades, 1982). Posteriormente, el NH_4^+ es desplazado a la solución por el K^+ y de esta forma puede obtenerse la capacidad de intercambio catiónico total del suelo.

Para la determinación de las bases de cambio se realizaron extracciones sucesivas con acetato amónico ($AcNH_4$) 1*M* ajustado a pH = 7. Se emplearon 10 g de suelo tamizado por 2 mm. Se les añadió 33 cm^3 de solución de $AcNH_4$ 1*M* y se agitó por volteo durante 5 min. Posteriormente se centrifugó a 2000 rpm durante 5 min, luego el sobrenadante se filtró y se recogió en un matraz de 100 cm^3. Este proceso se repitió un total de tres veces y los matraces donde se recogieron los sobrenadantes se enrasaron con la misma solución de $AcNH_4$ que se usó en las extracciones. La concentración de bases de cambio se determinó por espectroscopia de emisión de plasma con un equipo ICP Perkin Elmer modelo Optima 4300 DV. El residuo de suelo de las extracciones anteriores se lavó con 33 cm^3 etanol al 96% agitando durante 5 min, se centrifugó a 2000 rpm y se desechó el sobrenadante. Este proceso se repitió dos veces más. Finalmente, se realizaron extracciones sobre ese residuo con 33 cm^3 de solución de KCl 1*M* con intervalos de 5 min de agitación y centrifugación, de manera semejante a las extracciones previas de las bases de cambio. El sobrenadante se filtró y se recogió en un matraz de 100 cm^3. Se empleó un conductivímetro Thermo Orion modelo 720 con un electrodo de ion selectivo para determinar la concentración del NH_4^+ desplazado por el KCl, con lo que se obtuvo la capacidad de intercambio catiónico total. Todos los resultados de CEC y cationes de cambio se encuentran en la Tabla 4.

2.8 Separación de la materia orgánica libre y de los ácidos fúlvicos libres

La materia orgánica libre (FOM) suele estar compuesta principalmente por restos orgánicos frescos o con un escaso grado de descomposición que, además, no suelen estar fuertemente ligados a la fracción mineral del suelo. La separación de la FOM se llevó a cabo por flotación (Dabin, 1971), empleando una solución de H_3PO_4 2*M* con una densidad de 1.12 $g \cdot cm^{-3}$. Para ello, se partió de muestras de 50 g de suelo, a las que previamente se les había extraído la fracción lipídica, y se les añadió 100 cm^3 de H_3PO_4 2*M.* En los casos en que se producía efervescencia, debido a la presencia de carbonatos, se les añadió

una gota de octan-1-ol para evitar la formación de espuma. Posteriormente, la suspensión de suelo con el líquido denso se agitó durante 1 min con un agitador de varilla y se centrifugó a 2000 × g durante 5 min. La suspensión se filtró con filtros cónicos de papel, quedando la FOM retenida en el filtro. El proceso se repitió una vez más; la FOM fue lavada varias veces con agua hasta alcanzar el pH neutro, con la finalidad de eliminar el H_3PO_4 que pudiera quedar retenido. Finalmente la FOM fue secada en estufa a 60 °C y pesada. La solución filtrada de H_3PO_4 2*M*, que presentaba coloración amarillenta, se conservó ya que este contenía los denominados ácidos fúlvicos libres (FFA). Se midieron los volúmenes de los extractos de FFA y luego se tomaron alícuotas de 50 cm^3 de las que tenían un color más claro y 25 cm^3 para los extractos más oscuros. Las alícuotas se neutralizaron añadiendo NaOH 5*M* y seguidamente se llevaron hasta sequedad calentando en un baño a 50 °C para finalmente determinar su contenido en C mediante el método de oxidación con $K_2Cr_2O_7$ (Walkley & Black, 1934).

Figura 2.6. Separación densimétrica de la materia orgánica libre (FOM) y obtención de la fracción de ácidos fúlvicos libres (FFA), empleando H_3PO_4 2*M*.

2.9 Obtención de las fracciones húmicas del suelo

Una vez separadas la FOM y los FFA, se procede a separar las fracciones restantes de la materia orgánica del suelo. Estas serían el extracto húmico total (THE), compuesto por los HA y los FA, y finalmente la humina.

Para la extracción de las fracciones restantes se empleó la metodología descrita por Dabin (1971) y Duchaufour & Jacquin (1975), basada en extracciones con soluciones de pH alcalino. A los suelos libres de FOM y FFA

se les añadió 100 cm^3 de $Na_4P_2O_7$ 0.1*M* y se agitaron durante 12 h. Posteriormente se centrifugaron a 2000 × g durante 5 min y el sobrenadante se recogió. El proceso se volvió a repetir una vez más. Después de emplear la solución de $Na_4P_2O_7$ 0.1*M* se realizaron extracciones con NaOH 0.1*M*. Para ello, se añadieron 100 cm^3 de la solución NaOH 0.1*M* y se agitaron las suspensiones durante 12 h. En esta fase, el número de extracciones con NaOH que se emplearon fue dependiente de la coloración de los extractos: por ello se fueron tomando alícuotas de cada una de las extracciones para realizar un seguimiento del color. La extracción se dio por finalizada cuando se alcanzó una coloración escasa y constante. Una vez finalizadas las extracciones se midió el volumen total de extracto y se tomaron dos alícuotas de cada muestra para realizar las determinaciones de THE y HA. Para la determinación del THE se llevó a sequedad la alícuota de 10 cm^3 del extracto y se determinó su contenido en C mediante el método de oxidación por $K_2Cr_2O_7$ (Walkley & Black, 1934). Para determinar el HA se acidificó otra alícuota de THE añadiendo gotas de HCl 6*M*. Una vez que el medio está a pH ácido se observa como precipitan los HA, que posteriormente se centrifugan a 46000 × g durante 10 min desechando el líquido sobrenadante que estaría compuesto por los FA. El precipitado se vuelve a disolver en el mínimo volumen posible de NaOH 0.5*M.* Finalmente se lleva hasta sequedad y se determina el C del HA por el método de oxidación por $K_2Cr_2O_7$ (Walkley & Black, 1934). El contenido de FA se obtiene por diferencia entre los valores de THE y HA (FA = THE - HA). Los resultados del fraccionamiento se encuentran en la Tabla 5.

Figura 2.7. Extracción y purificación de los ácidos húmicos (HAs). Determinación del C del extracto húmico total (THE) y HA por valoración mediante el método de oxidación por $K_2Cr_2O_7$.

Una vez determinados el THE y los HAs se procedió a la separación preparativa y purificación de estos segundos. Para ello, se acidificó todo el extracto con HCl 6*M* para precipitar los HAs. A continuación, se sifonó la mayor cantidad de volumen posible sin generar pérdidas de precipitado. El volumen restante se separó del precipitado centrifugando a 2000 × g durante 5 min desechando el líquido sobrenadante. Una vez retirada toda la solución líquida que contenía los FA se volvió a disolver el precipitado en el menor volumen posible de NaOH 0.5*M*. Una vez disuelto el HA, se centrifugó a 46000 × g durante 15 min para que sedimentaran las arcillas y se recogió la solución de humato sódico libre de arcillas. Se volvió a precipitar la solución de HA empleando HCl 6*M*, que luego se centrifugó a 10000 × g durante 5 min desechando la fracción sobrenadante. Por último, las muestras se sometieron a diálisis empleando membranas Medicell® (size 6 Inf Dia 27/32”–21.5mm) para eliminar los cloruros. La diálisis se dio por concluida cuando el agua de la diálisis no mostraba presencia de Cl^- (test con $AgNO_3$), con lo que los HA se secaron a 40 °C y se pesaron. El análisis elemental de los HA purificados se llevó a cabo en un equipo LECO CHNS-932, el O se determinó por diferencia a partir de la muestra libre de cenizas. La Tabla 6 refleja todos los resultados del análisis elemental.

2.10 Caracterización de la materia orgánica del suelo

2.10.1 Espectroscopía visible

La presencia de grupos cromóforos y radicales libres estables en la estructura de las sustancias húmicas hace posible una primera caracterización de las mismas a través de la espectroscopía visible. Estos grupos son capaces de absorber radiación UV-visible, por lo que esta técnica constituye una herramienta rápida y sencilla para obtener una aproximación del grado de trasformación o madurez de los HAs toda vez que, por definición, se considera que el proceso de humificación se caracteriza por la formación progresiva de sustancias coloreadas no existentes en la biomasa original. Concretamente, dos longitudes de onda han sido tradicionalmente seleccionadas como índices de la intensidad del color de soluciones de HA ajustadas a una concentración constante; estas absorbancias son determinadas a 465 y 665 nm, y denominadas E4 y E6 respectivamente. Estas absorbancias, así como la relación E4/E6 son usadas como indicador del grado de humificación o transformación (Jiménez-Ballesta et al., 1983), si bien también han podido ser relacionadas con la aromaticidad de la estructura de los HA (Traina et al., 1990; Tinoco et al., 2015).

Cuando se obtiene la segunda derivada de los espectros de HAs, pueden reconocerse y cuantificarse algunos máximos de intensidad variable a 616 y 572 nm. Estos son considerados característicos de cromóforos polinucleares de tipo quinónico, que en el caso de los suelos son interpretados como indicadores de presencia de hongos (Kumada & Hurst, 1967; Sato & Kumada, 1967; Marynowski et al., 2015). La determinación de la densidad óptica a estas longitudes de onda se llevó a cabo empleando soluciones de HAs en NaOH 0.01*M* con concentración de 0.2 $mg \cdot cm^{-3}$ (Kononova, 1982) usando un espectrofotómetro Hewlett-Packard HP 8452A. La Tabla 6 muestra los resultados de espectroscpía visible.

2.10.2 Espectroscopía infrarroja

Esta técnica espectroscópica se basa en suministrar a la muestra energía comprendida en el intervalo infrarrojo del espectro electromagnético, con lo que

se produce la absorción de parte de esta energía, generandose una excitación vibracional en los enlaces de las moléculas, que experimentan distintos movimientos (tensiones y flexiones) dependiendo del tipo de enlace. El estudio estructural de los HA se lleva a cabo en la región del IR medio, comprendida entre 4000 y 400 cm^{-1}, aunque la mayor parte de bandas de compuestos orgánicos aparecen entre los 2000 y 600 cm^{-1} (a excepción de la banda característica a 3400 cm^{-1} de tensión de los grupos O–H y la de 2900 cm^{-1} correspondiente a la tensión de C–H alifático). Los modos de vibración de los enlaces tras la excitación se pueden clasificar en los siguientes tipos: movimientos de tensión (pueden ser simétricos o asimétricos), y modo de flexión de tipo tijera, balanceo, torsión y aleteo.

Para obtener los espectros de IR de los HA, se usaron 2 mg de HA y 200 mg de KBr. Se homogeneizaron en un mortero de ágata, manteniendo siempre las muestras libres de humedad, y se prensaron en forma de pastilla para la medida en el equipo. Se empleó un espectrofotómetro Agilent Cary 630 FTIR para la obtención de los espectros en un rango de 4000—400 cm^{-1} con una resolución de 4 cm^{-1}.

2.10.3 Espectroscopía de resonancia magnética nuclear de ^{13}C en estado sólido

Se trata de una técnica que proporciona valiosa información estructural de la materia orgánica y unos valores semicuantitativos de la proporción de las distintas estructuras. Se trata de una técnica no destructiva, por lo que se puede recuperar la muestra tras su análisis. La resolución de los espectros en estado sólido no es tan elevada como en las muestras en solución, pero al estar la muestra más concentrada, los tiempos de adquisición del espectro se reducen sustancialmente. El tiempo de adquisición de los espectros y la calidad de los mismos depende de la abundancia de los átomos con núcleos con número impar de protones o neutrones (en este caso el ^{13}C), que interaccionan con un campo magnético externo, emitiendo una señal de radiofrecuencia. Las regiones del espectro de NMR, características de las distintas estructuras, han sido descritas por varios autores (González-Vila et al., 1983; Knicker, 2011): alquil + *α*-amino C (0–45 ppm); *N*-alquil + OCH_3 C (45–60 ppm); *O*-alquil C (60–110 ppm); C aromático no sustituido (110–140 ppm); C aromático

heterosustituido (140–160); C en grupos carbonilo (principalmente carboxilo + amida, 160–220 ppm).

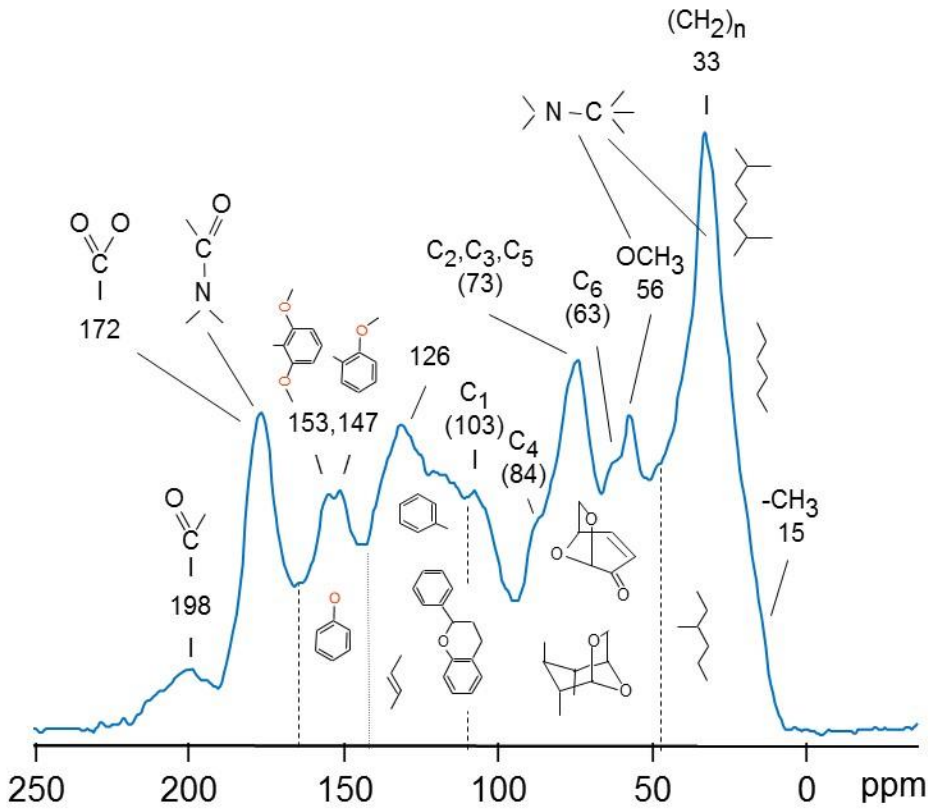


Figura 2.8. Espectro de resonancia magnética nuclear (NMR) de ^{13}C en estado sólido de un ácido húmico (HA). Se muestra la asignación de las principales regiones y picos (Almendros, 2008). En este trabajo se usaron las regiones: alquil + *α*-amino C (0–45 ppm); *N*-alquil + OCH_3 C (45–60 ppm); *O*-alquil C (60–110 ppm); C aromático no sustituido (110–140 ppm); C aromático heterosustituido (140–160); carbonil C (principalmente carboxilo + amida, 160–220 ppm).

Los espectros fueron obtenidos en un instrumento Bruker modelo Avance 400 MHz a una frecuencia de 100.63 MHz utilizando rotores de óxido de circonio de 4 mm de diámetro y tapas de Kel-F®. Se usó la polarización cruzada (CP) durante la rotación en el ángulo mágico (MAS) con una frecuencia de rotación de 12.5 kHz. Se acumularon entre 5000 y 6000 scans con un retardo de pulso de 300 ms. Se aplicó una rampa de 1 ms de tiempo de contacto para evitar la interacción dipolar heteronuclear de Hartmann-Hahn. Los desplazamientos químicos del ^{13}C se calibraron usando tetrametilsilano (0 ppm). La distribución cuantitativa de los distintos tipos de C se determinó por integración de las

distintas regiones usando el software Mestrenova 10 (Mestrelab Research). Los resultados de la integración de las distintas regiones del NMR se muestran en la Tabla 7.

2.10.4 Pirólisis acoplada a cromatografía de gases/ espectrometría de masas (Py-GC/MS)

Esta técnica es una de las más versátiles para la caracterización de la materia orgánica, por cuanto no precisa aplicar pretratamientos químicos. Permite el análisis de muestras sólidas de forma directa en un horno pirolizador que se encuentra acoplado al inyector cromatográfico. La muestra sólida experimenta un rápido calentamiento generando compuestos de pirólisis volátiles, que se hacen pasar por la columna cromatográfica. Finalmente, estas moléculas son fragmentadas en un detector de espetrometía de masas y los iones generados son separados y cuantificados utilizando como detector un espectrómetro de masas cuadrupolar. A través de la abundancia de estos fragmentos y el uso de librerías establecidas se pueden identificar los distintos compuestos en el cromatograma. Se trata de una técnica destructiva, ya que la muestra es pirolizada en el microhorno, pero presenta la ventaja de que la cantidad empleada es muy baja (1–2 mg). Con esta técnica se llevó a cabo la caracterización de las muestras de suelo completo para obtener una idea semicuantitativa de las proporciones de compuestos en la materia orgánica.

Se utilizó un pirolizador modelo PY-2020iD (Frontier Lab Ltd., Fukushima, Japan) acoplado a un cromatógrafo de gases (GC/MS) Agilent modelo 6890. La columna empleada fue de fenilmetilsiloxano (Agilent HP-5MS 5%). La pirolisis se llevó a cabo con un calentamiento a 500 °C durante 1 min. La fase móvil utilizada fue He con un flujo de 1 $cm^3 \cdot min^{-1}$. La temperatura del horno era de 50 °C durante 1 min, luego se incrementó hasta 100 °C con una velocidad de 30 $°C \cdot min^{-1}$, y finalmente hasta 300 °C con una rampa de calentamiento de 10 $°C \cdot min^{-1}$. Posteriormente se mantuvo a temperatura constante de 300 °C durante 10 min. El detector de masas era un cuadrupolo Agilent 5973, y los espectros fueron adquiridos usando una energía de ionización de 70 eV. Para la identificación de los picos se emplearon las librerías NIST y Wiley.

2.10.5 Espectrometría de masas de alta resolución por resonancia ciclotrónica de ión

Se trata de una técnica de alta resolución para el estudio de masas moleculares de compuestos volátiles generados tras la inyección de una solución en un sistema de electrospray. Las moléculas son ionizadas y confinadas en un campo magnético de alta intensidad. Una vez atrapados los iones, estos se excitan y adquieren un movimiento circular en un dispositivo de tipo ciclotrón, debido al campo magnético. Los iones giran de acuerdo a su relación masa/carga y posteriormente son detectados. La resolución en este tipo de medidas es mayor cuanto mayor es el campo magnético utilizado. Esta técnica fue empleada en el presente trabajo para llevar a cabo un estudio de la estructura de los HAs a nivel molecular con alta resolución.

Figura 2.9. Equipo Bruker 12 Tesla Apex-Qe de espectrometría de masas de alta resolución por resonancia ciclotrónica de ión (FTICR-MS) equipado con una fuente de ionización de electrospray (ESI) Apollo II (Old Dominion University, Norfolk, EEUU).

Para este análisis se empleó un instrumento Bruker 12 Tesla Apex-Qe FTICR-MS equipado con una fuente de ionización de electrospray (ESI) Apollo II (Fig.

2.9). Se tomaron 0.5 mg de HA purificado y se disolvieron en la mínima cantidad necesaria de NH_4OH al 1%. Previamente a la inyección, las muestras fueron diluidas con una mezcla de metanol:agua (1:1, v:v) con el fin de mejorar la eficiencia de la ionización. Se asignaron las formulas empíricas a las moléculas detectadas en el rango de valores 200 a 800 m/z, para ello se realizó una calibración empleando ácidos grasos naturales equivalentes a los presentes en las muestras (Sleighter et al., 2008).

2.11 Análisis de datos

2.11.1 Análisis discriminante

Con este tipo de análisis estadístico se pretende establecer, de acuerdo a unos criterios preestablecidos, si una muestra o individuo pertenece a un grupo u otro. Se trata de un análisis descriptivo. En este tipo de análisis los grupos están definidos desde un principio, lo cual lo diferencia del análisis de cluster o de conglomerados. Se trata de evaluar a qué grupo pertenece, o debería pertenecer cada individuo. El hecho de obtener una buena clasificación de muestras indicaría que las variables independientes empleadas para el análisis, y los grupos definidos para la clasificación guardan una relación.

2.11.2 Regresión de mínimos cuadrados parciales

La regresión de mínimos cuadrados parciales (PLS) es un análisis estadístico que reduce los predictores a un conjunto más pequeño de componentes no correlacionados y realiza una regresión de mínimos cuadrados sobre estos componentes, en lugar de hacerlo sobre los datos originales. La PLS se utiliza para encontrar las relaciones fundamentales entre dos grupos de datos, es decir, un enfoque de variable latente para modelar la estructura de covarianza en estos dos espacios. El modelo obtenido como resultado de PLS trata de explicar la variación de un grupo o variable (variable dependiente) de acuerdo con otro grupo (variables independientes), así como la posible relación existente entre ellos.

Para la evaluación y selección de un buen modelo se usan diversos índices. Entre ellos se encuentra el criterio de Akaike (1974) o la raíz del error cuadrático

medio (RMSE). Aparte de ello, es conveniente repetir el análisis usando valores aleatorizados de la variable dependiente (variable que se quiere predecir) con objeto de verificar que el modelo no está sobreajustado.

La obtención de un modelo significativo indicaría relación entre las variables predictoras y las predichas, además el valor obtenido de la importancia de las variables para la predicción (VIP) daría idea de las variables predictoras que aportan más información al modelo predictivo.

2.11.3 Análisis de componentes principales

El análisis de componentes principales (PCA) es un tipo de análisis estadístico en el cual se intenta describir un conjunto de datos en términos de unas nuevas variables denominadas “componentes” que no se encuentran correlacionadas entre sí. Se trata de un analsis de carácter exploratorio para observar relaciones entre variables. Se intenta representar una serie de muestras y variables que pueden estar correlacionadas en un plano descrito por componentes que no se encuentran correlacionadas.

Además, se pueden introducir variables suplementarias; estas variables serian variables que no han sido usadas para calcular las componentes pero se podrían representar en el plano factorial obtenido de tal manera que ayudaran a la interpretación de los resultados.

2.11.4 Escalado multidimensional

El análisis de escalado multidimensional (MDS) constituye una técnica multivariante para la visualización y exploración de datos. En un espacio definido por un número seleccionado de dimensiones se representan, en este caso, las distintas variables, que se situan de tal manera que las distancias entre ellas son proporcionales a la afinidad entre las mismas, definida por el índice de 1-Pearson de correlación. La configuración final de variables se lleva a cabo mediante un algoritmo iterativo que parte de una configuración de puntos aleatoria, y se detiene al alcanzar el mejor ajuste posible entre las distancias en el plano y el índice de similaridad utilizado, todo ello expresado mediante un parámetro denominado “stress” que es la una suma residual de cuadrados (Kruskal, 1964).

Capítulo 3:

Soil carbon storage predicted from the diversity of pyrolytic alkanes

Este capítulo ha dado lugar a la publicación: Jiménez-González M.A., Álvarez A.M., Hernández Z., Almendros G., 2018. Soil carbon storage predicted from the diversity of pyrolytic alkanes. Biology and Fertility of Soils 54:617–629.

https://doi.org/10.1007/s00374-018-1285-6

Abstract

Biogeochemical factors responsible of the highly variable content of soil organic matter (SOM) in the different types of soils are poorly known. In particular, the role of organo-mineral interactions has frequently been considered, but less attention has been paid to the molecular composition of the SOM. The aim of this work was to contribute to a better qualitative and quantitative assessment of the soil organic C (SOC) accumulation, using chemometric approaches that do not require the absolute knowledge of the structure and functioning of the whole system under study. For this reason, we monitored the *n*-alkanes released by analytical pyrolysis from 35 widely different Mediterranean soils. The *H′* Shannon diversity index was calculated to evaluate the origin and transformations of the alkane homologous series (C_9–C_{31}). A series of multivariate data treatments succeeded in showing significant relationship between the diversity of alkanes and the SOC concentration, and additional indicators of SOM quality were also used. All statistical analyses pointed out the significant correlation ($P < 0.01$) between the *H′* diversity of the pyrolytic alkanes and the amount of SOC. In particular, a significant relationship between SOC levels and the percentage of long-chain alkanes was found, whereas the percentage of short-chain alkanes was correlated with specific descriptors of SOM quality. Finally, the partial least squares (PLS) predicted the SOC content from the alkane patterns.

3.1 Introduction

The biogeochemical factors involved in the structural transformations of soil organic matter (SOM) are highly variable in the spatial and temporal scales, and for this reason, soils have different concentrations of soil organic C (SOC). Therefore, accurate monitoring of soil C storage processes is crucial for a quantitative assessment of the global C cycle, because the total amount of organic C stored in soil not only is an important C pool excluded from the atmospheric circulation (Lal, 2004) but also depends on the stabilized, humic-type organic matter (Kögel-Knabner et al., 1988). For these reasons, in recent decades, much research has focused on modeling the C storage of different soils. Generally, the stabilization of SOM has been simulated by simple models mainly considering SOC preservation through organo-mineral interactions (Schmidt et al., 2011) and met the molecular structure of SOM and the specific enzyme processes. However, in vitro incubation experiments have showed a large variability in the rates of transformation of plant and microbial residues into stabilized humic-type material (Almendros & Dorado, 1999; Pendall & King, 2007; Requena et al., 1996). Other research showed how microbial and fungal community to large extent depends on the composition of different SOM fractions (Montiel-Rozas et al., 2017; Tian et al., 2017).

The aliphatic hydrocarbon content of the SOM can give information on organisms and processes involved in SOC storage (Almendros et al., 1991; Baldock et al., 1992; Kögel-Knabner and Hatcher, 1989; Nip et al., 1986). Saturated lipids and especially alkanes are more refractory to microbial oxidation than unsaturated lipids (Gocke et al., 2013). Apart from the environmental information provided by the alkane signature of the different soils, these compounds also play a role affecting different soil physical properties. Their hydrophobic character can improve soil aggregation (Dinel et al., 1991; Jambu et al., 1983, 1985) or promote soil water repellency (Jiménez-Morillo et al., 2016; Jordán et al., 2013; Rumpel et al., 2004). By means of physicochemical interactions (encapsulation, solid solution, diffusion into microporous structures, hydrophobic bonding, etc.), alkanes can also condense into the structure of humic substances, accumulating in the course of SOC storage (Amblès et al., 1991; Eglinton and Logan, 1991; Nebbioso et al., 2015). For these reasons, the environmental information provided by the different alkyl compounds in soil can be considered (Jansen & Wiesenberg, 2017).

The main goal of this work was to contribute to a better understanding of the mechanisms of SOC sequestration in wide variety of soils differing mainly in terms of vegetation, geological substrate, and land use, using chemometric assessments that do not require the detailed knowledge of the structure of the whole system under study. A series of indices more or less responsive to SOC content were compared. Therefore, a chemical and statistical assessment was carried out, considering the alkanes released by analytical pyrolysis as indicators of the SOC content.

3.2 Materials and Methods

3.2.1 Study area

Thirty-five soil samples at (0–10 cm) were surface collected from forest and agricultural ecosystems (Table 1) so as to reflect the wide variability in soil types. The soils differed mainly in terms of geological substrate and vegetation. After removing the litter layer, composite soil samples were prepared mixing soil material collected from three sampling points; then, the soil samples were air-dried and sieved (< 2 mm), and all visible plant fragments were removed during sieving.

3.2.2 Soil physical and chemical analysis

We classified soils according to the IUSS Working Group WRB (2014) system (Table 1). Soil texture was determined by using the densimeter method (Bouyoucos, 1927). Soil pH was measured in soil-water suspension (1:2.5, w:w) using an XS pH meter model pH 7 (Carpi, MO, Italy). The electrical conductivity (EC) was measured with an XS COND7 conductivity meter (Carpi, MO, Italy) using soil-water slurry (1:5, w:w). Soil bulk density was determined using a cylindrical core of known volume (Blake & Hartge, 1986). The cation exchange capacity (CEC) was measured with ammonium acetate according to Juo et al. (1976). The SOC concentration was determined by wet chemical oxidation with 1*N* potassium dichromate (Nelson & Sommers, 1982; Walkley & Black, 1934). The N was determined by micro-Kjeldahl digestion (Prince, 1945). The water holding capacity (WHC) of soils was determined in the laboratory at atmospheric

pressure (Klute, 1986). A series of routine variables indicating the quality of the SOM were determined. In particular, different SOM fractions were isolated: free organic matter (FOM), fulvic acid (FA), humic acid (HA), and humin, in order to assess the transformation extent of the SOM. The FOM was removed by flotation with 2*M* H_3PO_4 (Dabin, 1976). The extractable SOM was extracted by 0.1*M* $Na_4P_2O_7$ and 0.1 *M* NaOH (Duchaufour & Jacquin, 1975). The HA and FA were separated by precipitating the alkaline soil extract with 6*M* HCl. The optical density (E4) of HA was determined at 465 nm at a concentration of 0.2 mg cm^{-3} in 0.01*M* NaOH (Kononova, 1966). In addition, absorbance values at 616 and 572 nm in the second derivative spectrum were also measured so as to determine the concentration of fungal chromophors ascribed to 4,9-dihydroxyperylene-3,10-quinone (DHPQ) structures (Kumada & Hurst, 1967; Sato & Kumada, 1967). Finally, solid-state ^{13}C NMR spectra of HA were integrated to obtain semi-quantitative data for the different SOC forms. The spectra were obtained in a Bruker Avance 400 MHz, operating at a frequency of 100.63 MHz with zirconium rotors of 4 mm OD with Kel-F caps. The cross polarization (CP) was used during magic-angle spinning (MAS) of the rotor at 12.5 kHz. Between 5000 and 6000 scans were accumulated with a pulse delay of 300 ms. A ramped 1H pulse was applied during the 1-ms contact time to circumvent Hartmann-Hahn mismatches. The ^{13}C chemical shifts were calibrated using tetramethylsilane (0 ppm). The spectral ranges were chosen as already reported (González-Vila et al., 1983; Knicker, 2011): alkyl + α-amino C (0–45 ppm); *N*-alkyl + OCH_3 C (45–60 ppm); *O*-alkyl C (60–110 ppm); aromatic C (110–160 ppm); carbonyl C, mainly carboxyl + amide (160–220 ppm). The ^{13}C intensity distribution was determined by integrating signal intensity over the above-mentioned chemical shift regions using MestreNova 10 software suite (Mestrelab Research).

3.2.3 Pyrolysis

Whole soil samples were analyzed by pyrolysis-gas chromatography-mass spectrometry (Py-GC/MS), which can determine the relative percentages of the different SOM components and does not require time-consuming extractions, fractionation, or sample pretreatments (De la Rosa et al., 2008; Jiménez-González et al., 2016).

Soil (2 g) was ground to fine powder (< 0.01 mm) using a planetary agate mill and analyzed by a PY-2020iD device (Frontier Lab Ltd., Fukushima, Japan) connected to an Agilent 6890 GC/MS system with a phenylmethylsiloxane column (Agilent HP-5MS 5%). Pyrolysis temperature was 500 °C for 1 min. Helium at a flow rate of 1 cm^3 min^{-1} was used as the carrier gas. The GC oven temperature was 50 °C for 1 min; then, it was increased to 100 °C at 30 °C min^{-1}, and from 100 to 300 °C at a rate of 10 °C min^{-1}, and kept isothermal at 300 °C for 10 min. The compounds were identified from their mass spectra acquired at 70 eV ionizing energy with an Agilent 5973 quadrupole mass spectrometer. The ion 85 trace was used to identify and quantify the alkanes. The identity of the alkanes was confirmed by their retention times (RTs), compared with those of standards, and with RTs of other pyrolytic alkyl compounds (C_{16} and C_{18} fatty acids with a diagnostic peak at m/z 60). The carbon preference index (CPI) (calculated as ratio between odd C and even C-numbered alkanes) and the ratio of short- to long-chain *n*-alkanes already used (Duan & He, 2011) were calculated, to get insight on the origin and transformation degree of alkanes in the different environments.

3.2.4 Statistical analysis

The variability of the concentration of SOC in the different soils (dependent variable) was described as a function of the composition of pyrolytic alkanes (independent variables) normalized as total abundances (sum of peak areas of the 23 *n*-alkanes = 100). The data were processed using either simple linear regression models or multivariate data treatments for exploring correlations between soil properties. Several indices to express diversity of the pyrolytic molecular assemblages were examined (Simpson, Margelef, and Shannon-Wiener), but the classical Shannon-Wiener diversity index (*H'*) was selected because of its simplicity. The percentages of the 23 *n*-alkanes were used to calculate *H'* using the Species Diversity & Richness II software version 2.5 (Pisces Conservation). The diversity was also independently calculated for different groups of alkanes: total alkanes, short- and long-chain alkanes (< C_{20} and > C_{20}, respectively), and odd and even C-numbered alkanes ($2n+1$ and $2n$, respectively).

The correlations between the SOC content and *H'* diversity indices were examined. In the second stage, principal component analysis (PCA) was carried

out using the total abundances of pyrolytic alkanes and the Statistica software ver. 7.1. The PCA results were shown as a biplot where soils are displayed as points while variables (the different alkanes) are displayed as vectors in new axes represented by principal components that explain the most variability between the variables. This multivariate data treatment was used with the aim of assessing the different contributions of individual alkanes to the total variability of SOC and SOM pool content. For this purpose, and in addition to alkanes, several indicators of SOM quality, such as E4, CEC, DHPQ, *H′* index, and ^{13}C NMR regions of HA, were processed as supplementary variables, i.e., those which are not used to calculate the ordination axes. The principal components were also calculated only using the alkane percentages as principal variables; the supplementary variables are represented in the space according to the components calculated from the principal variables (Legendre & Legendre, 1998).

Partial least squares regression (PLS) was applied to simulate models to explain the maximum percentage of the variance of the SOC content. PLS was carried out with the Statgraphics Centurion XV software, using the total abundance of *n*-alkanes as independent variables and the SOC content as dependent variable. To select the minimum number of latent variables (LVs) of each PLS model ($P < 0.05$) the criteria considered during the cross-validation test set with the leave-one-out method were the average prediction sum of squares (PRESS) and the root mean square error (RMSE). In addition, a more rigorous criterion was used, consisting of repeating PLS using fully randomized SOC values as dependent variables (the models were discarded as overfitted if some significant ($P < 0.05$) model with random dependent variable was obtained).

Multidimensional scaling (MDS) used for variables classification was applied to explore mutual relationships between alkane abundances and other soil dependent variables, including the SOC values and using the 1-Pearson *r* correlation index as a measure of similarity. The closeness in the plane between variable scores indicates their mutual correlations. This may be used for unsupervised classification or clustering of the different variables (independent variables: alkane abundances, dependent variables: SOM descriptors). The MDS was carried out using the Statistica software version 7.1.

3.3 Results

3.3.1 Semi-quantitative and qualitative assessment of the SOM in the soils

The soil samples presented a large variability of SOC content (from 17 to 157 $g \cdot kg^{-1}$) as well as other soil properties and the SOM fractions (Tables 5, 6, 7, 8). As it concerns the SOM fractions (FOM, FA, and HA), they showed a large variability in amount and quality. In particular, the HA showed different values of E4 and in ^{13}C NMR spectra. This was also the case with the content of DHPQ, which may be an indicator of microbial (fungal) presence in soil (Kumada & Hurst, 1967; Sato & Kumada, 1967; Marynowski et al., 2015).

3.3.2 Alkane homologous series

The relative abundances of the alkanes, which have been obtained using the ion trace 85 m/z, are shown in Fig. 3.1. The series ranged between nonane (C_9) and hentriacontane (C_{31}), and generally odd C-numbered alkanes were more abundant than the even C-numbered alkanes (CPI > 1). The ratio of short-to-long-chain *n*-alkanes values is also shown in Table 8. Generally, long-chain alkanes with strong predominance of components with an odd number of C atoms (Eglinton & Hamilton, 1967) are characteristic of higher plants, whereas in microbial communities and lower plants, this predominance decreases or it does not exist (Stránský et al., 1967).

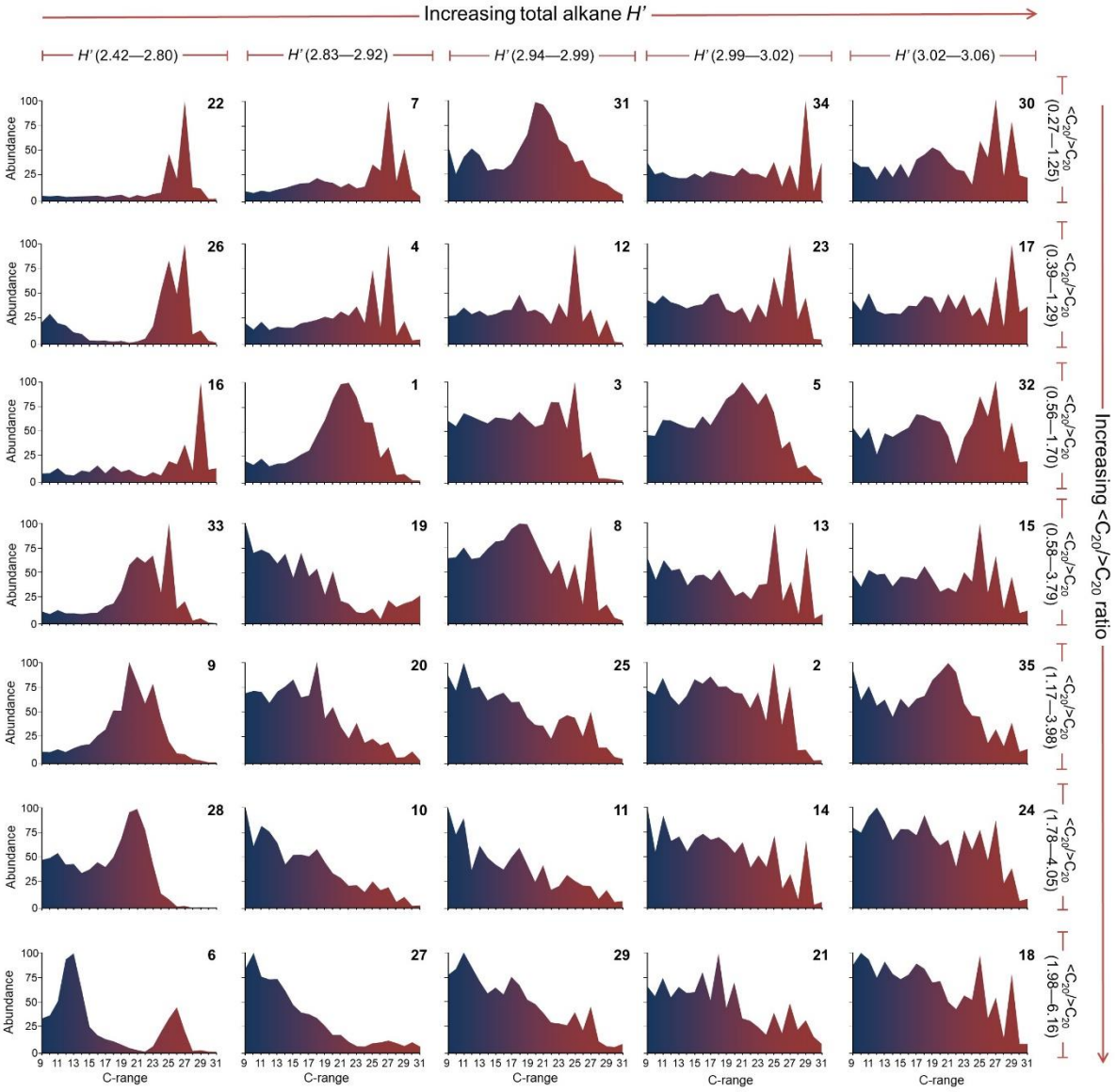


Figure 3.1. Alkane homologous series released by pyrolysis from whole soil samples. Series are ordered horizontally in function of increasing diversity and vertically in function of increasing values of the $<C_{20}$ / $>C_{20}$ ratio. Numbers of the series refer to Table 1.

3.3.3 Principal component analysis

Figure 3.2 shows the biplot obtained by PCA with the two first principal components calculated from the abundances of pyrolytic alkanes, and including a series of soil properties as supplementary variables mainly related with SOC content and quality, as well as the diversity indices obtained from the various alkane subsets. This plot explains 37.6% (first component) and 27.3% (second component) of the total variability, respectively.

Figure 3.2 shows that the alkane eigenvectors define three independent clusters represented by (i) short-chain alkanes (C_9–C_{18}); (ii) medium-chain alkanes from C_{19}–C_{23}; and (iii) long-chain alkanes (C_{24}–C_{31}). The short-chain alkane cluster shows a sharp independent subgroup formed by C_9–C_{14} alkanes. As regards the supplementary variables, the SOC content and the different alkane *H′* diversity indices were closest to each other in the plot. Therefore, diversity indices have greater correlation with SOC than the other tested supplementary variables. On the other hand, the eigenvectors of variables related to the quality of SOM are direct in the opposite direction (Arom C, E4), coinciding to medium-chain length alkanes.

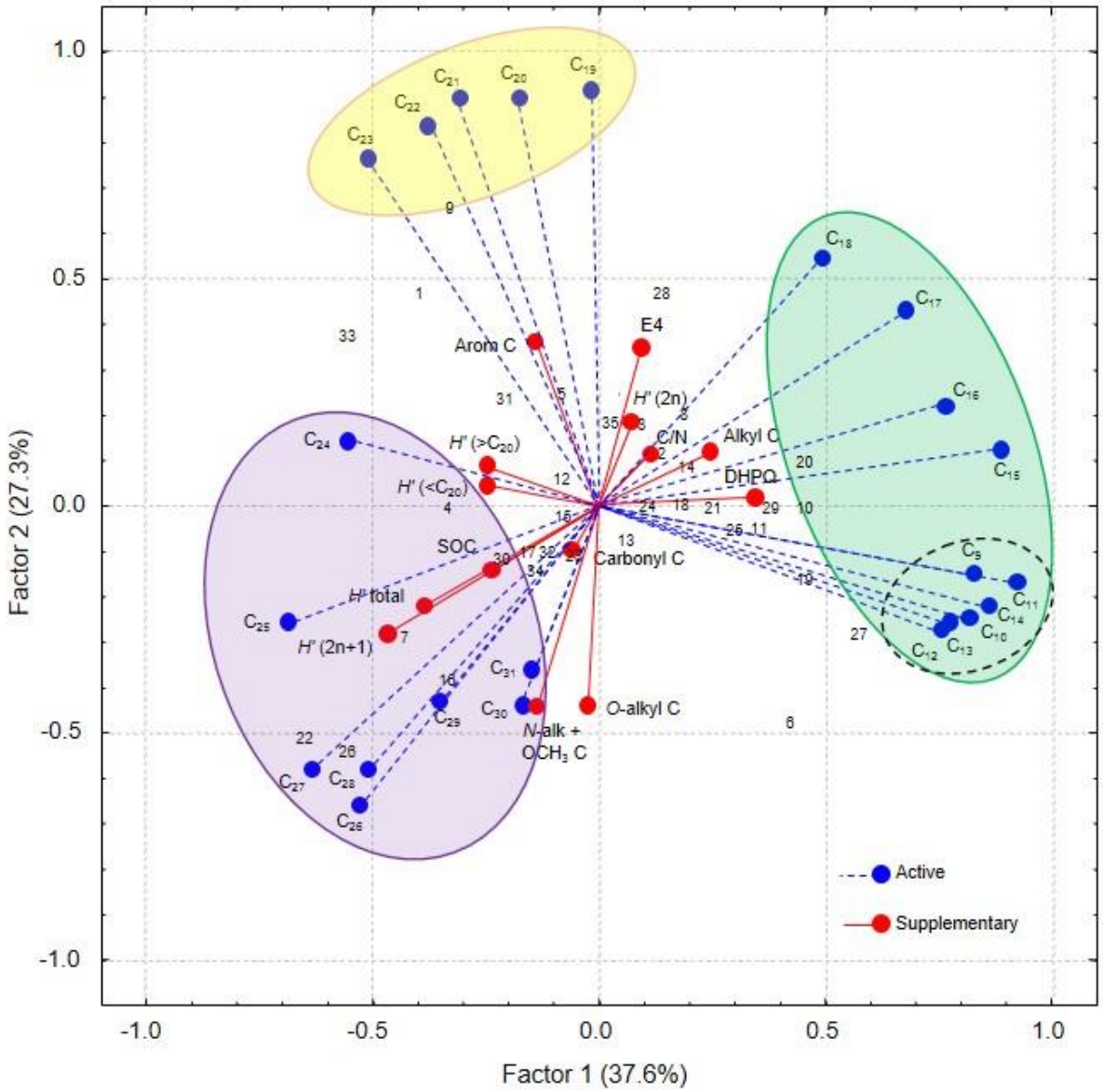


Figure 3.2. Biplot obtained by principal component analysis, with numbers representing soil samples and eigenvectors pointing to the direction of increasing values of the independent variables (alkane abundances, represented with blue dashed lines). A series of SOM properties (considered as dependent variables and processed as supplementary variables) are displayed with solid red lines. The plot shows a pattern for alkanes suggesting four clusters for the independent variables. Variable labels refer to Tables 4, 7 and 8: soil organic C (SOC), optical density of HA at 465 nm (E4), peak intensity at 616 and 572 nm in the 2nd derivative visible spectra of the HAs (DHPQ), different regions of ^{13}C NMR (Arom C, Alkyl C, Carbonyl C, *O*-alkyl C and *N*-alk + OCH_3 C), and Shannon index (*H'*) of short-chain, long-chain, odd, and even alkanes (< C_{20}, >C_{20}, 2*n*+1 and 2*n*, respectively).

3.3.4 Linear regression models

The Shannon diversity index *H′* that was independently calculated for each homologous series of alkanes (Table 8) displayed a large variability within the soil samples. Whereas the diversity index *H′* that was calculated from the total alkane series showed a significant correlation with the SOC, different correlations were obtained for *H′* calculated from a narrower alkane series, i.e., only of alkanes with specific C number or chain length (Fig. 3.3). For instance, the *H′* calculated from the total alkane series showed a very significant correlation with the SOC at *P* value = 0.0015, R^2 = 0.2655, whereas the *H′* calculated from the odd C-numbered alkanes (*P* value = 0.0012, R^2 = 0.2756) showed better correlations than those for long- and short-chain alkanes and for even C-numbered alkanes (*P* value = 0.0272, R^2 = 0.1394; *P* value = 0.0125, R^2 = 0.1745; and *P* value = 0.0136, R^2 = 0.1709, respectively) although in all cases the correlations were significant at $P < 0.05$.

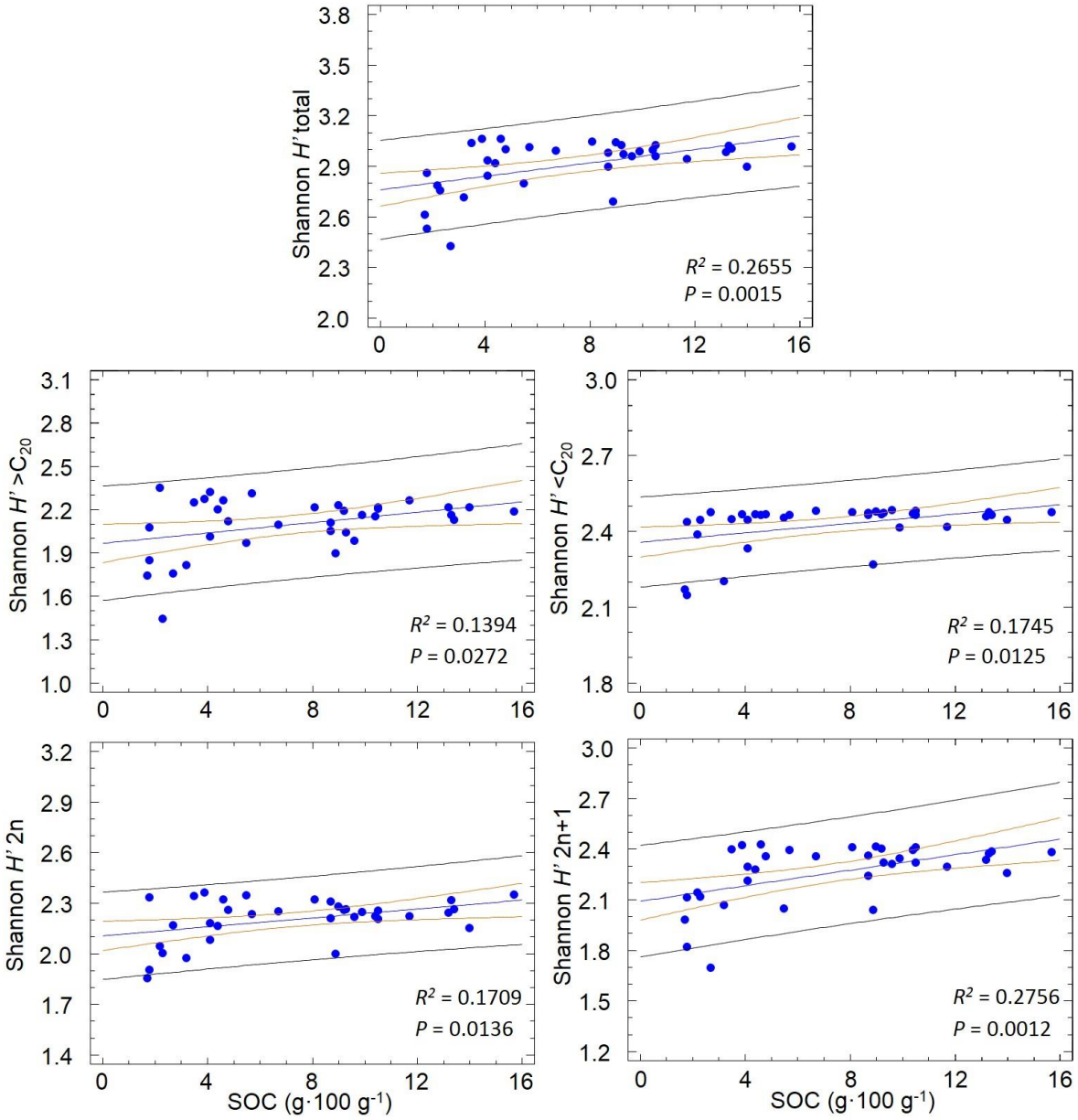


Figure 3.3. Comparison of the linear relationships between the concentration of soil organic C (SOC) and the Shannon diversity index (*H'*) calculated from total alkanes and from the short-chain (< C_{20}), long-chain (> C_{20}), even C-numbered (2*n*), and odd C numbered (2*n*+1) alkanes released by pyrolysis from whole soil samples. Inner (orange) and outer (black) lines indicate 95% confidence and prediction limits, respectively.

3.3.5 Partial least squares prediction models for soil C

The results obtained by PLS are in line with the previous correlations. PLS gave a highly significant ($P < 0.01$) forecasting model for predicting SOC concentration (Fig. 3.4a). The lack of overfitting in this model was confirmed by repeating the PLS model using randomized values of the dependent variable SOC. If the values of SOC that are not real (randomized) gave a non-significant model and the real values gave a significant one, it means that the prediction model is not based in spurious information. In all cases, with the same number of LVs (15), the random SOC values led to a poor correlation ($P > 0.05$) in the cross-validation plot (Fig. 3.4b). Figure 3.4c shows the coefficient plot and the cross-validation plot for the experimental SOC values, which correspond to the coefficients for each alkane in the equation of the prediction model generated by PLS.

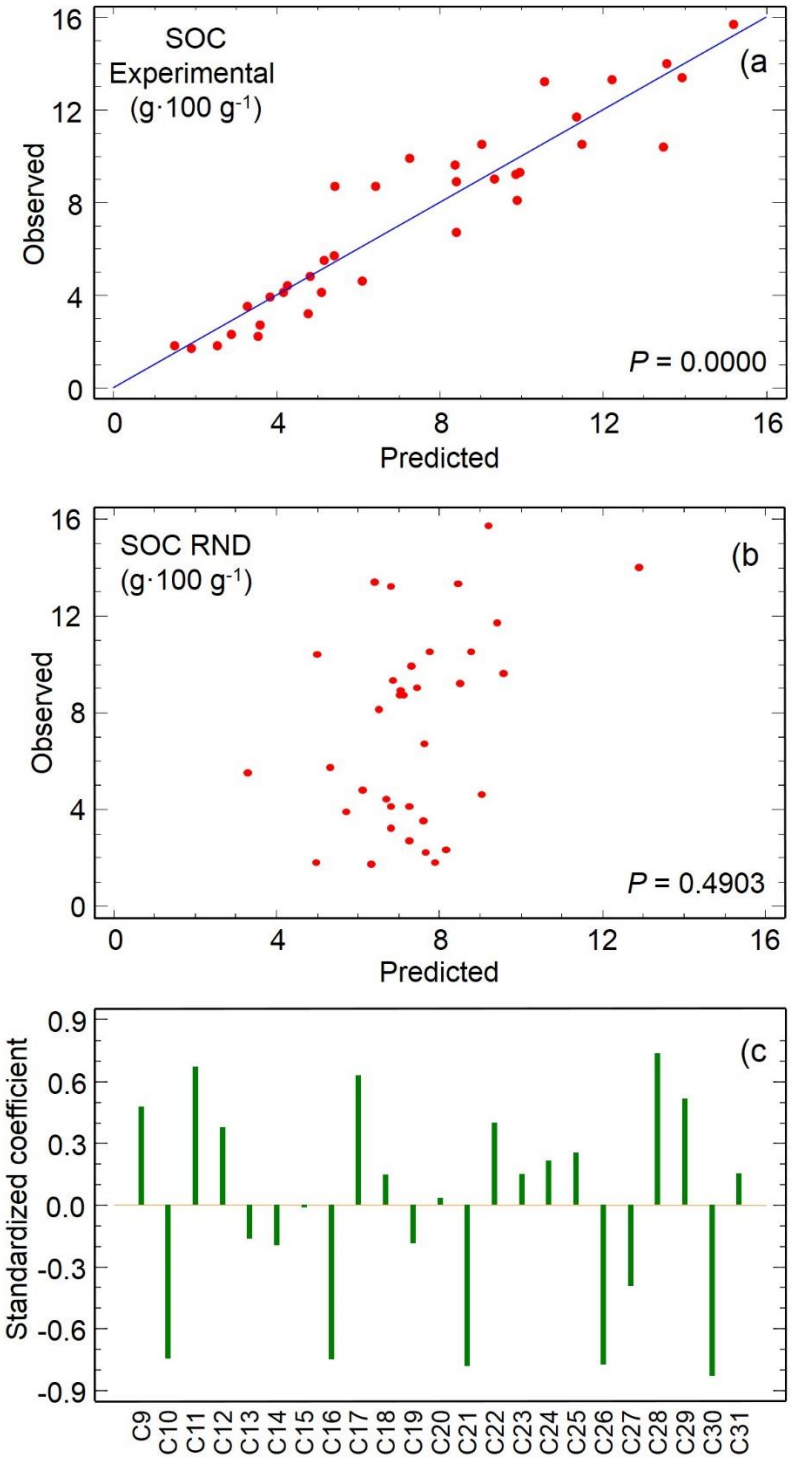


Figure 3.4. Observed versus predicted values of soil organic C (SOC) content calculated by partial least squares (PLS) regression using pyrolytic alkanes from whole soil samples as descriptors to predict: **a)** experimental SOC values, **b)** fully randomized SOC values. Standardized coefficients of the 23 *n*-alkanes corresponding to the PLS forecasting model for the SOC are shown in **c)**.

3.3.6 Multidimensional scaling

The results of the MDS, used to classify soil variables, are presented as a scatter diagram (Fig. 3.5), where the final stress was 0.2011, indicating excellent reliability of the scatterplot (Kruskal, 1964). The final configuration defined by the variable scores after MDS clearly shows that the five *H'* diversity indices are the variables that explain to the largest extent the SOC values. Other variables (FOM, WHC, and, to lesser extent, soil C/N ratio) were also clustering close to SOC and diversity indices. On the other hand, soil properties related to SOM quality such as E4, Arom C, and carbonyl C tended to form a cluster in the opposite side of the scatterplot.

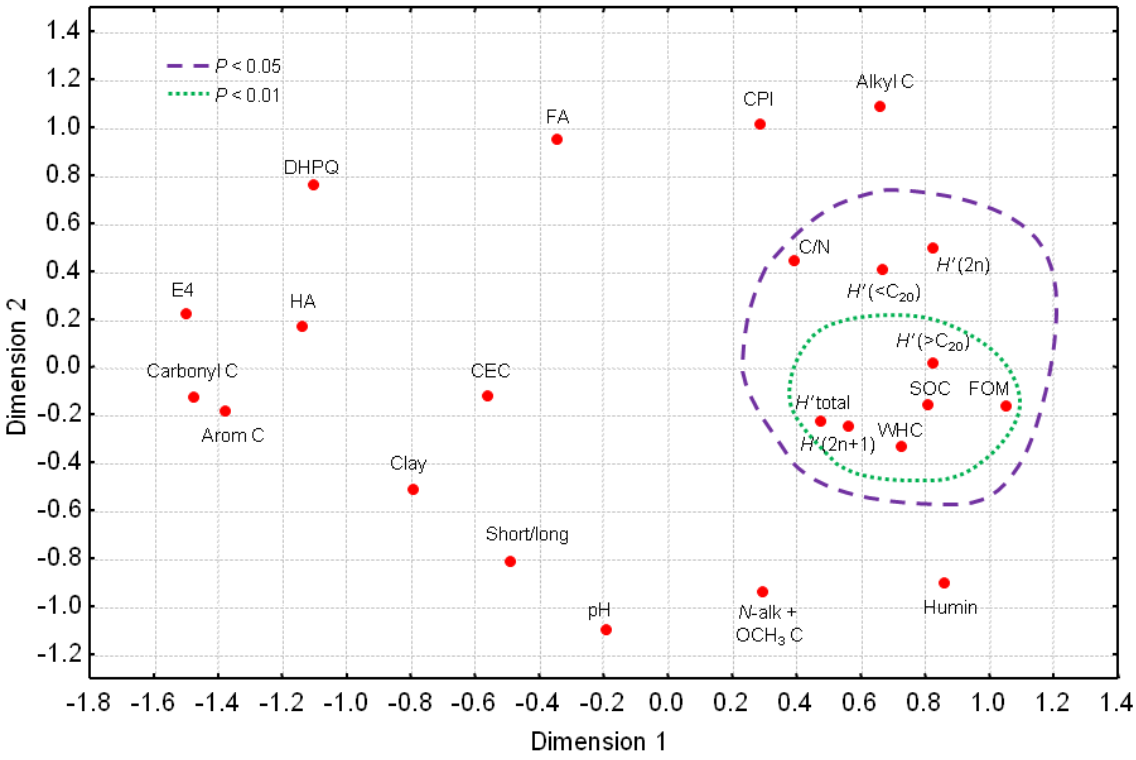


Figure 3.5. Representation, using multidimensional scaling (MDS), of the relationships between different soil properties and quality descriptors of SOM, and the Shannon *H'* diversity index calculated from different groups of pyrolytic alkanes. Concentric lines indicate the variables positively correlated with the SOC at $P < 0.01$ (dotted line) and at $P < 0.05$ (dashed line). Variable labels refer to Tables 3, 4, 6, 7 and 8.

3.4. Discussion

The fact that the diversity index calculated from pyrolytic *n*-alkanes showed the correlation with the SOC content may suggest a relationship between the complexity of the molecular composition of the SOM and its content in soil. This confirms previous studies, suggesting different SOM molecular structures in relation to the levels of organic C in the different soils (Almendros et al., 2018; Jiménez-González et al., 2017). The increase in alkane diversity may be related to the resistance to the biodegradation of the SOM. In soils where SOM is more effectively preserved, there are more chances of accumulation of complex molecular compounds derived from organisms inhabiting soil. Banerjee et al. (2016) suggested that the extent of SOM transformation in arable soils depends on the redundancy of bacterial and fungal communities, and Wu et al. (2010) found a significant correlation between SOC concentration and the diversity of the corresponding vegetation.

The independent PLS analysis showed that the total abundance of pyrolytic alkanes, used as independent variables, significantly predicted ($P < 0.01$) the SOC content (Fig. 3.4), probably because the pyrolytic alkanes accumulated information on the sources and aging of SOM. This memory of the alkane signature is probably enhanced by an increasing complexity of soil components and biogeochemical processes. In fact, pyrolytic alkanes, but not free lipid alkanes, are largely entrapped within micro-aggregates and humic-type substances and thus physically protected against the degradation in soil. Therefore, pyrolytic alkanes act as a molecular record of past organisms and processes.

Different degradation techniques showed that, in the case of alkanes, pyrolysis is superior to mild wet chemical degradation in providing unbiased information (Tinoco et al., 2002), because pyrolysis caused an early thermoevaporation of alkanes, whereas chemical degradation led to continuous alteration of labile aliphatic molecules. This smoothes the alkane biogenic signature and severely modifies the indices (CPI, average chain length (ACL), etc) that are used to define the extent of the biogeochemical transformations.

The representation of the eigenvalues of the independent variables (alkane abundances) in the space defined by the two first principal components showed different clusters. These groups can be assigned to the origin of the

corresponding alkanes. Long-chain alkanes (> C_{20}) may derive from epicuticular waxes of vascular plants whereas short-chain alkanes (< C_{20}) may have a microbial origin (Simoneit & Mazurek, 1982). The small cluster formed by C_9–C_{14} alkanes may derive from thermal cracking produced by the pyrolysis. In Fig. 3.2, the supplementary variables SOC and *H′* indexes (of total and odd C-numbered alkanes) are grouped close to the long-chain alkanes and this may suggest the presence of fresh SOM, as supported by the concentration of specific alkanes with the characteristic signature of waxes from higher plants. On the other hand, variables indicating SOM quality or maturity such as CEC, E4, Arom C, or DHPQ tended to cluster together with medium-chain alkanes of probable microbial origin. In conclusion, despite the effectiveness of pyrolysis in breaking macromolecular structures of SOM, the alkane patterns still exhibit a characteristic biogenic signature, indicating that probably thermal cracking did not occur, and most alkanes were released in the early thermoevaporation stages.

The fact that the alkanes *H′* were the variables most closely related to the concentration of SOC is confirmed by the groups formed by MDS. In addition, this analysis showed that the WHC was situated near the SOC probably because retention of water in soil favors the accumulation of SOM, but because the increase in SOC content can improve soil structure and water retention (Nath, 2014). The same occurred with the FOM, indicating that fresh organic matter content prevails in soils with high SOC levels. Also in the MDS analysis, indicators of SOM quality or maturity were grouped in an isolated cluster. These indicators were the E4 and the aromaticity (suggesting condensed C structures), the C=O groups, the HA content, and the concentration of DHPQ (indicative of microbial accumulation of HA-like fungal melanins), which were located in the opposed side of the plot. As a whole, the MDS plot suggests a comparatively higher quality of the humic substances in soils with higher than lower biological activity and this may lead to active mineralization of the SOM.

3.5 Conclusions

The Shannon *H′* diversity index of alkanes released by analytical pyrolysis shows a very significant correlation with the total SOC stored in soil. This would suggest that the diversity of alkanes may indicate the complexity of the

biogeochemical processes in addition to the functional redundancy of the soil system. This makes possible the prediction of the SOC content from the signature of pyrolytic alkanes using PLS. The information supplied by the pyrolytic alkanes differs mainly in terms of their chain length, which may indicate a different origin. The long-chain alkanes (C_{24}–C_{31}), which are more correlated with fresh organic matter than other alkanes, showed the best significant correlations with the SOC levels. On the other hand, the *H′* indices calculated from the medium-chain alkanes correlated better than those calculated from other alkanes with soil factors related with the SOM quality and maturity. Finally, the very short-chain alkanes (C_9–C_{18}) were unrelated to SOC content, possibly due to their origin from pyrolytic cracking of longer alkanes or condensed SOM alkyl structures.

Capítulo 4:

The diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of soil carbon storage

Este capítulo ha dado lugar a la publicación: Jiménez-González M.A., Álvarez A.M., Carral P., González-Vila F.J., Almendros G., 2017. The diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of soil carbon storage. Journal of Chromatography A 1508:130–137.

https://doi.org/10.1016/j.chroma.2017.05.068

Abstract

The variable extent to which environmental factors are involved in soil carbon storage is currently a subject of controversy. In fact, justifying why some soils accumulate more organic matter than others is not trivial. Some abiotic factors such as organo-mineral associations have classically been invoked as the main drivers for soil C stabilization. However, in this research indirect evidences based on correlations between soil C storage and compositional descriptors of the soil organic matter are presented. It is assumed that the intrinsic structure of soil organic matter should have a bearing in the soil carbon storage. This is examined here by focusing on the methoxyphenols released by direct pyrolysis from a wide variety of topsoil samples from continental Mediterranean ecosystems from Spain with different properties and carbon content. Methoxyphenols are typical signature compounds presumptively informing on the occurrence and degree of alteration of lignin in soils. The methoxyphenol assemblages (12 major guaiacyl- and syringyl-type compounds) were analyzed by pyrolysis-gas chromatography-mass spectrometry. The Shannon-Wiener diversity index was chosen to describe the complexity of this phenolic signature. A series of exploratory statistical analyses (simple regression, partial least squares regression, multidimensional scaling) were applied to analyze the relationships existing between chemical and spectroscopic characteristics and the carbon content in the soils. These treatments coincided in pointing out that significant correlations exist between the progressive molecular diversity of the methoxyphenol assemblages and the concentration of organic carbon stored in the corresponding soils. This potential of the diversity in the phenolic signature as a surrogate index of the carbon storage in soils is tentatively interpreted as the accumulation of plant macromolecules altered into microbially reworked structures not readily recognized by soil enzymes. From a quantitative viewpoint, the partial least squares regression models exclusively based on total abundances of the 12 major methoxyphenols were especially successful in forecasting soil carbon storage.

4.1 Introduction

Despite the considerable progress achieved during the few past years in the knowledge of the factors presumptively involved in soil carbon storage, additional research is required to establish the variability of these factors in the space and time (Swift, 2001; Yuan et al., 2012; Wenting et al., 2014). Such knowledge is crucial to progress in the development of the scientific bases of Earth's biogeochemical cycle and the impact of the invoked global change. Most studies center their attention on the organo-mineral interactions (Lützow et al., 2006; Kögel-Knabner et al., 2008; Hernández et al., 2012; Wenting et al., 2014), whereas comparatively fewer researches have considered the possible role of the intrinsic macromolecular complexity of the soil organic matter (SOM) on the soil carbon storage. The lack of statistically-validated information of the variable role of the factors responsible of SOM stabilization in different ecosystems encourages the interest of further exploratory research looking for useful biogeochemical proxies, informing on the extent to different non-excluding factors may be active in the soil carbon storage process.

The SOM fraction presents heterogeneous structure in part derived from microbial transformation of plant biomacromolecules. Assuming the large variability of the chemical composition of the SOM, most research has focused on the analysis of specific molecular features providing environmental information on the mechanisms and processes involved in the formation of the SOM fractions (Hernández et al., 2012), which should be closely related to the performance of soil C storage. In particular, analytical tools such as pyrolysis-gas chromatography-mass spectrometry (Py-GC/MS) have lead to substantial progress not only in fingerprinting SOM structure in terms of their origin, but also in releasing molecules with high diagnostic value as biomarker compounds. In fact, recent research on the methoxyphenol composition of the fraction of humic acid (HA) in soil (Miralles et al., 2015) have suggested compositional pyrolytic descriptors informative about C stabilization processes in local scenarios.

Methoxyphenols have an origin from plant biomacromolecules mainly lignin, and this family of compounds represents typical pyrolysis products from SOM (Martín et al., 1979). Unlike other plant or microbial-derived biomacromolecules which turn upon pyrolysis into secondary diagnostic products (e.g., carbohydrates leading to cyclic ketones or furans), in the case of lignins the

structure of the original building blocks (guaiacyl (G), syringyl (S) and *p*-OH phenyl building blocks) are preserved to large extent.

The vegetation type in the different soils is important in the resulting methoxyphenol signature. In fact, classical studies have established the differences existing between the lignin from gymnosperms, angiosperms or grass plants in terms of the phenolic composition (Tinoco et al., 2002). Major methoxyphenols released by pyrolysis from soil organic matter are 4-H-, methyl-, ethyl-, vinyl-, propenyl- and aceto- derivatives of guaiacol and syringol (Tinoco et al., 2002) although it is well known that the monomer composition of lignins highly varies depending on the botanical source (Fengel & Wegener, 1984). Syringol derivatives are practically lacking in pyrolysates from soils under gymnosperm plants while both S and G units are present in angiosperm plants.

On the other hand, the relative amount of G or S-type structures is very important in the biodegradability of lignin. The presence of the second methoxyl group at S-lignin units is associated to a more linear and less condensed structure. For this reason, S-lignins confer resistance with flexibility to plants, which can be important to herbaceous plants or plant organs that must grow quickly (Bonawitz & Chapple, 2010). Accordingly, the use of the S/G ratio as an index for lignin degradation is based on the idea that G-type lignins are more resistant to biodegradation due to its condensed structure.

In the present research, we study the methoxyphenol signature in the SOM, in a wide variety of continental Mediterranean soils from Spain, in order to explore the presumable relationships between its composition, defined as total abundances of the different types and the resulting molecular diversity, and the performance of the soil carbon storage in the corresponding soils as reflected by multivariate chemometric models.

4.2 Materials and Methods

4.2.1 Soils studied

Up to 35 topsoil samples (0–10 cm) were collected from Spanish soils with contrasted values of carbon content, geological substrate, vegetation and

microclimate. Three sampling points were taken in each sampling area (10 m^2). After removing the litter layer, soil materials from three sampling points were pooled to prepare the composite soil sample. Then these composites were air-dried and homogenized by sieving (fine earth, < 2 mm).

4.2.2 Physical and chemical properties of the soils

The texture was determined following the Bouyoucos method (Bouyoucos, 1927). Soil pH was measured in soil-water suspension (1:2.5), using a pH-meter pH7 (XS Instruments). A similar procedure was followed for the electrical conductivity (EC) but using soil-water slurry (1:5) and a COND7 (XS instruments). The cation exchange capacity (CEC) was measured with ammonium acetate according to Juo et al., (1976). The soil organic carbon (SOC) was measured by the Walkley and Black wet oxidation method (Walkley & Black, 1934; Nelson & Sommers, 1982). Total N was determined by micro-Kjeldahl digestion (Prince, 1945). Soils classification (Table 1) was carried out according to FAO (2014).

The main chemical and spectroscopic variables of the SOM from the samples were determined aiming to identify descriptors showing correlation with SOM quantity or its quality. After removing of free organic matter (FOM) by flotation with H_3PO_4, the extractable soil organic matter was isolated with standard method (Duchaufour & Jacquin, 1975) using 0.1*M* sodium pyrophosphate and sodium hydroxide. HA and fulvic acid (FA) were separated by precipitation with 6*M* HCl. For optical density measurements, solutions of HA in 0.1*M* NaOH were prepared with a concentration of 0.2 mg cm^{-3} (Kononova, 1982) to record the visible spectra and the absorbance at 465 nm (E4). The solid-state ^{13}C NMR spectra of HA were obtained in a Bruker Avance III HD 400 MHz instrument operating at a frequency of 100.64 MHz and using ZrO_2 rotors of 4 mm O.D. with Kel-F caps. The ^{13}C chemical shifts were calibrated relative to tetramethylsilane (0 ppm). The spectra were quantified by subdividing them into the following chemical shift regions: alkyl + *α*-amino C (0–45 ppm); *N*-alkyl + methoxyl C (45–60 ppm); *O*-alkyl C (60–110 ppm); aromatic C (110–160 ppm); carbonyl C (mainly carboxyl + amide 160–220 ppm). The ^{13}C intensity distribution was determined by integrating signal intensity over the above-mentioned chemical shift regions using the MestreNova 10 software.

4.2.3 Analytical pyrolysis

The SOM composition (whole soil samples) was studied by Py-GC/MS, which is a technique that displays large potential to assess the relative proportions of the different soil organic components (De la Rosa et al., 2008a). To study the soil samples by analytical pyrolysis, subsamples of 2 g were reduced to fine powder (< 0.01 mm) with a planetary ball mill with agate grinding bowls.

The pyrolysis was carried out with a double shot pyrolyser PY-2020iD (Frontier Lab Ltd., Fukushima, Japan) coupled to a GC/MS system Agilent 6890 equipped with a phenylmethylsiloxane column (Agilent HP-5MS 5%). Pyrolysis was conducted at 500 °C for 1 min. The carrier gas was helium at a flow rate of 1 cm^3 min^{-1}. The GC oven temperature was held at 50 °C for 1 min, then increased to 100 °C at 30 °C min^{-1}, from 100 °C to 300 °C at 10 °C min^{-1}, and isothermal at 300 °C for 10 min. The detector consisted of an Agilent 5973 quadrupole mass spectrometer, and mass spectra were acquired at 70 eV ionizing energy. Additional single ion chromatographic traces were extracted for the different compounds in order to quantify the series of homologues or the individual compounds as total abundances. The NIST and Wiley libraries were used to confirm the identity of the 12 signature methoxyphenols: guaiacol (G), methylguaiacol (MG), ethylguaiacol (EG), vinylguaiacol (VG), propenylguaiacol (PG), acetylguaiacol (AG), syringol (S), methylsyringol (MS), ethylsyringol (ES), vinylsyringol (VS), propenylsyringol (PS) and acetylsyringol (AS).

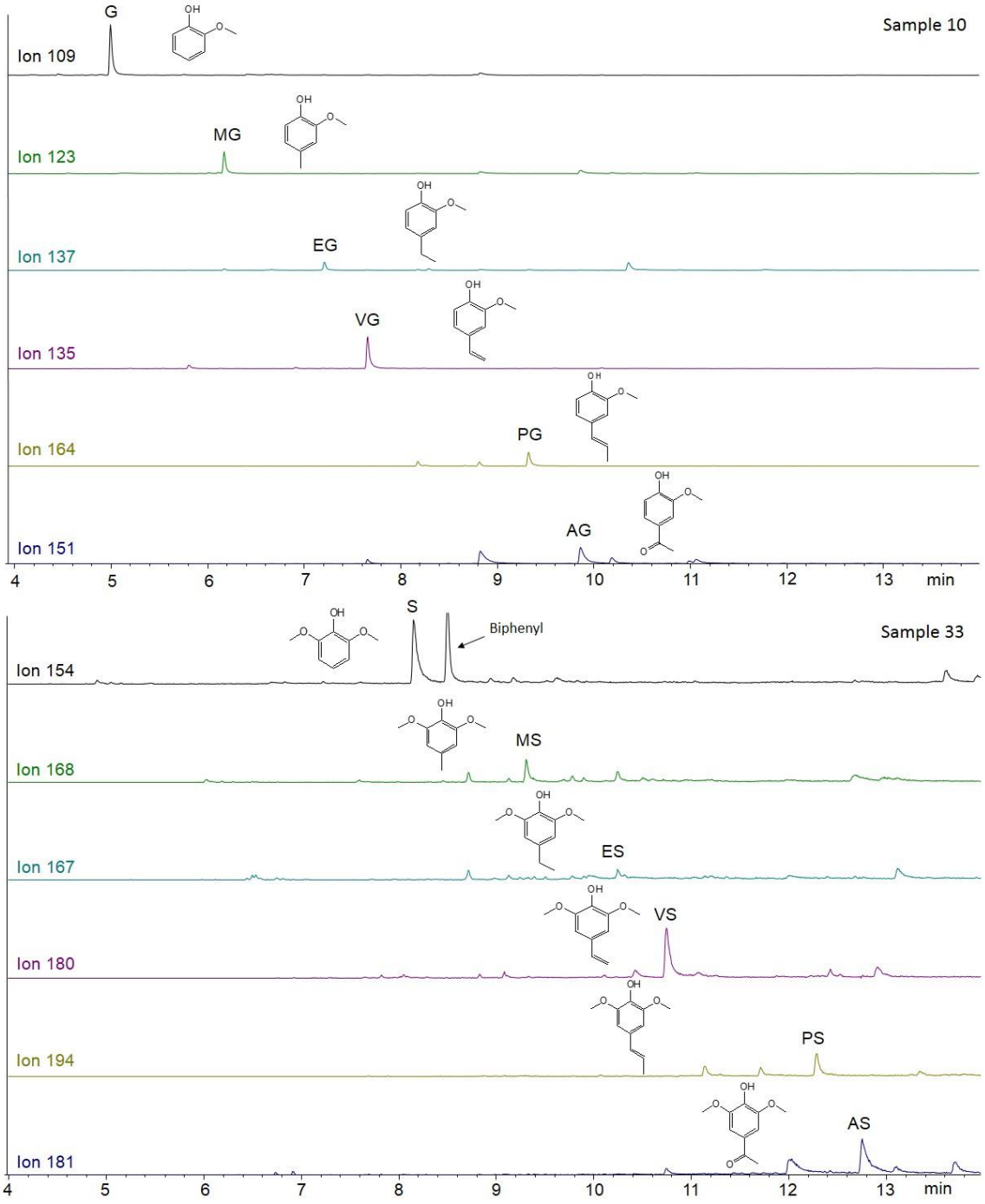


Figure 4.1. Chromatographic traces for diagnostic ions of the major methoxyphenols released by analytical pyrolysis from soil organic matter (SOM) (guaiacol (G), methylguaiacol (MG), ethylguaiacol (EG), vinylguaiacol (VG), propenylguaiacol (PG), acetylguaiacol (AG), syringol (S), methylsyringol (MS), ethylsyringol (ES), vinylsyringol (VS), propenylsyringol (PS) and acetylsyringol (AS)).

4.2.4 Data analysis

Several multivariate data treatments were used to explain the origin of the variability of the concentration of SOM in the soils (dependent variable) as a function of the composition of the major pyrolysis products (independent variables). We applied the Shannon-Wiener diversity index (*H'*), classically used to characterize species diversity in a community. In this case, the 12 methoxyphenols were selected as 'species' and their total abundances were used to calculate this index with the software Species Diversity & Richness II version 2.5.

Simple regression analysis was used to compare the *H'* biodiversity with the SOC in different soil samples and in sample subsets. Multivariate data treatments were carried out with the software Statgraphics Centurion XV, using the total abundances of 12 methoxyphenols after square root transformation as independent variables. Partial least squares regression (PLS) was applied considering the 12 methoxyphenols as independent variables and the SOC content as dependent variable. Spurious models due to overfitting were detected and discarded after repeating PLS models with fully randomized dependent variables. Discriminant analysis was also performed to check for the extent to which the methoxyphenol composition of the SOM may represent a source of proxies for the prediction of vegetation types. Finally, multidimensional scaling (MDS) using the software Statistica ver. 7.1 was used for simultaneous ordination of different soil dependent and independent variables, illustrating their mutual relationships as well as with the SOC levels.

4.3 Results

The general properties of the soils (Tables 2, 3 and 4) illustrate the large variability in the SOC (from 17 to 157 $g \cdot kg^{-1}$). One out of the sources of variability is the vegetation type, which in most cases also depends of the soil use. To analyze the response to vegetation with the methoxyphenol signature, two principal groups of vegetation were considered (angiosperms, gymnosperms) and a third, comparatively small group with herbaceous species referred to as the grass group (Table 2).

The 12 methoxyphenols identified by Py-GC/MS and their diagnostic ion traces are shown in Fig. 4.1. The peak close to S corresponds to biphenyl which also has a characteristic ion at 154 m/z.

The *H'* diversity index calculated from the total abundance of methoxyphenols (Table 8) showed large variability within the samples. When correlating SOC with *H'* separately calculated for the S and G families (Fig. 4.2), a highly significant correlation was observed, i.e., the increase in the diversity (or complexity) of the methoxyphenol assemblages is systematically accompanied by an increase in the SOC in the corresponding soils with a statistically significant level (*P* value = 0.0005 for the G group and *P* value = 0.0008 for the S group).

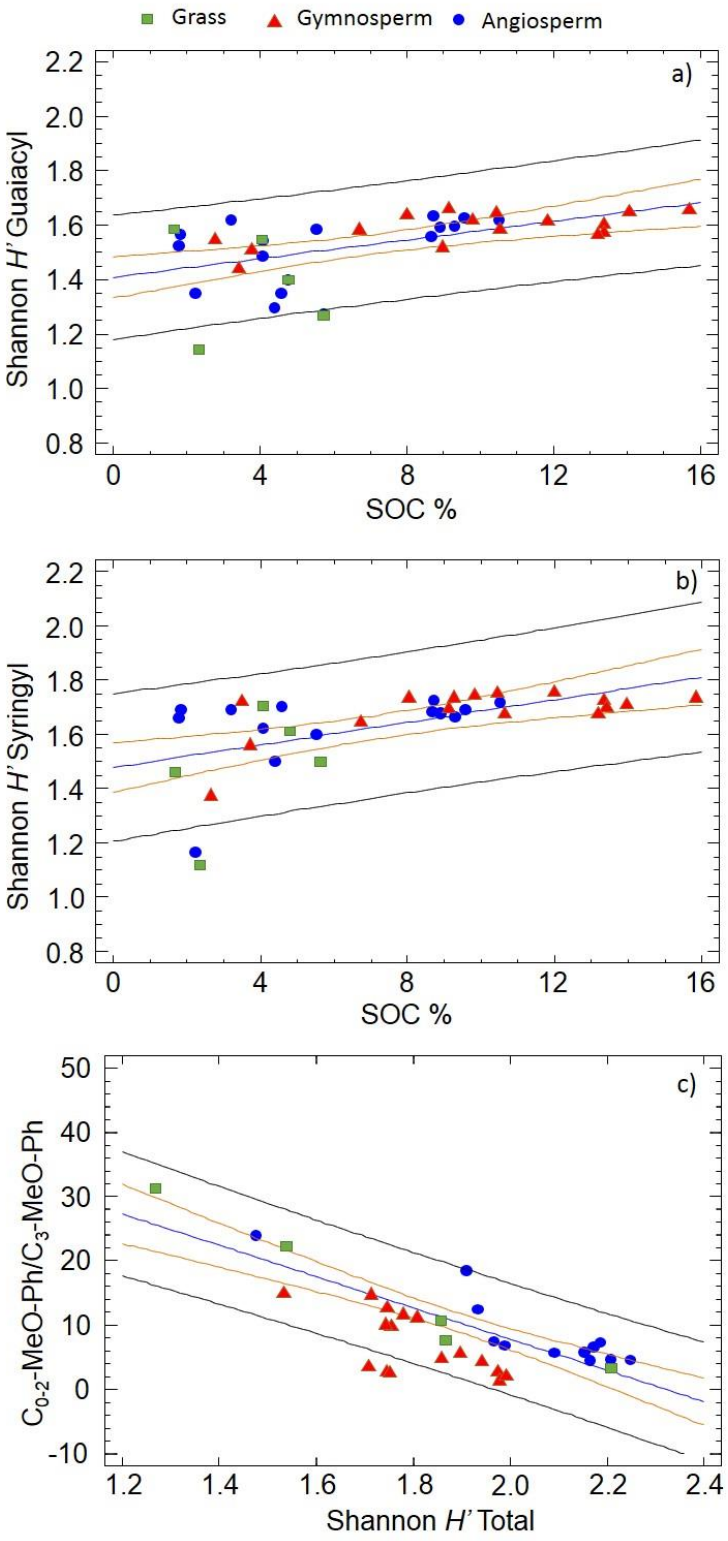


Figure 4.2. a, b) Relationship between the concentration of soil organic carbon (SOC) and the Shannon's biodiversity index (*H'*) calculated from the 12 major methoxyphenols released from soils by Py-GC/MS and c) biogenic methoxyphenol signature expressed as ($C_{0–2}$-MeO-Ph/C_3-MeO-Ph), with Shannon's biodiversity index. Inner and outer lines indicate 95% confidence and prediction limits, respectively.

A series of biodiversity indexes calculated from the 12 methoxyphenols (mainly Shannon-Wiener index, Simpson index, Margalef index, equitability and Berger-Parker index) were also calculated (data not shown), leading to similar conclusions and small differences as regards the significance levels. Consequently, the Shannon's *H'* was selected as being the most frequently used, not intending further justification of the different features of biodiversity which are selectively explained for these alternative indices.

When checking whether *H'* is correlated with other pyrolytic indices informing on changes in the structure of lignin, a poor correlation with the S/G ratio ($P > 0.05$) was found, but a very significant correlation exists with the ratio $C_{0–2}$-MeO-Ph/C_3-MeO-Ph ($P < 0.000$). This index (methoxyphenols with less than 3 alkyl C/(PG + PS) would inform on the degradation and oxidation of the C_3 side-chain in the structural units of lignin and suberin macromolecules (Fig. 4.2), which can be considered as a reflection of the progressive loss with decomposition of the biogenic lignin signature.

The PLS exclusively using total abundances of methoxyphenols as independent variables and including the first four PLS components led to highly significant ($P < 0.01$) forecasting models explaining the concentration of organic C in the different soils (Fig. 4.3). The validity of the models was confirmed by comparison with the alternative models computed from the fully randomized SOC (values of SOC assigned to the soil that not correspond with the real value in the soil) as dependent variable, which had a poor correlation ($P > 0.05$) with the total abundances of the methoxyphenols, whereas the real SOC values resulted in an excellent correlation with the methoxyphenol composition. In Fig. 4.3 (cross-validation plot) the observed values are represented versus the predicted values. The standardized coefficients for the methoxyphenols, obtained by PLS, are also shown in Fig. 4.3.

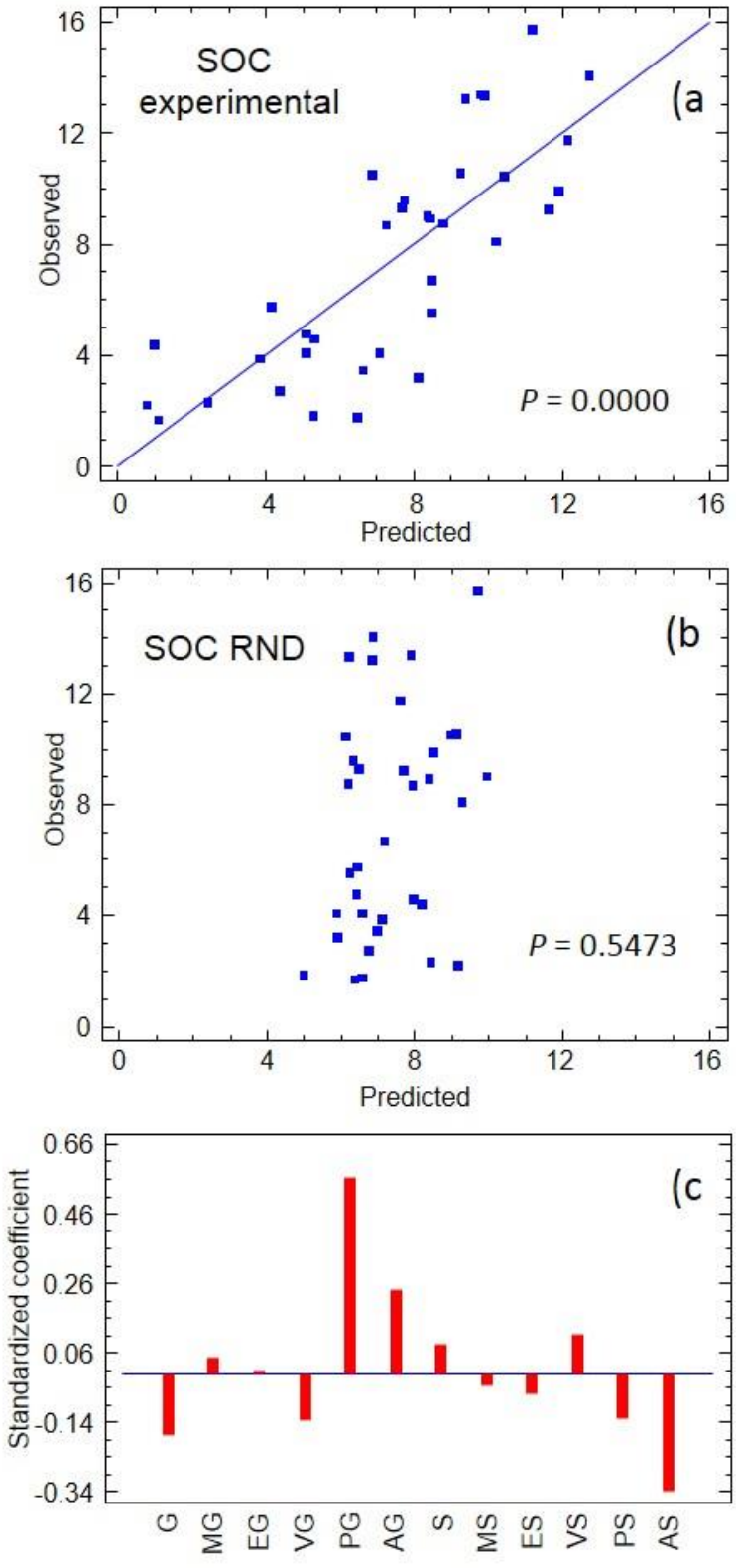


Figure 4.3. Observed versus predicted values of soil organic carbon (SOC) content calculated by partial least squares regression (PLS) from a) experimental values, b) fully randomized SOC values; c) standardized coefficients of 12 methoxyphenols obtained for SOC in the PLS forecasting model.

The results obtained in the discriminant analysis show an outstanding prediction potential of the methoxyphenols as regards vegetation type (gymnosperms, angiosperms or grasses). In Fig. 4.4 it is possible to observe the soil samples grouped in three sharp clusters. The three groups of vegetation in the 35 samples include 97% of cases correctly classified.

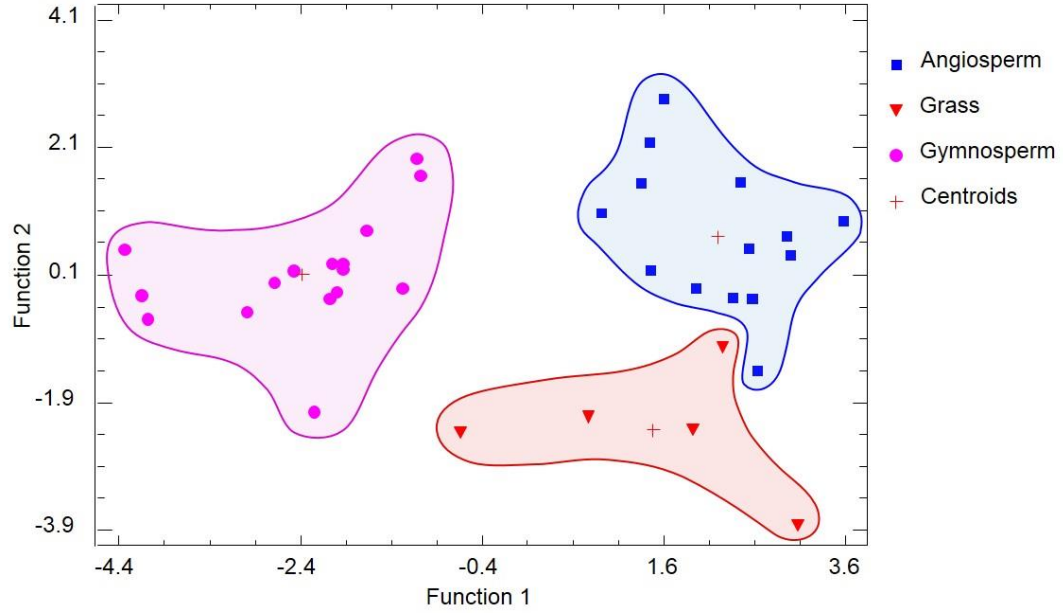


Figure 4.4. Discriminant analysis using the 12 pyrolytic methoxyphenols as variables and the vegetation type (angiosperm, gymnosperm and grass) as classification factor.

The results of MDS analysis are represented as a scatter diagram in Fig. 4.5. In this plot the scores for the different variables are proportional to the 1-Pearson indices from the correlation matrix including dependent and independent variables. In this plot the variables represented by points more close in the plane also correspond to those with more significant correlation indices between. The 'stress' level in the diagram was 0.2169, indicating excellent reliability of the scatterplot (Kruskal, 1964) (i.e., low spatial distortion when the original *n*-dimensional space is represented in the 2D plane).

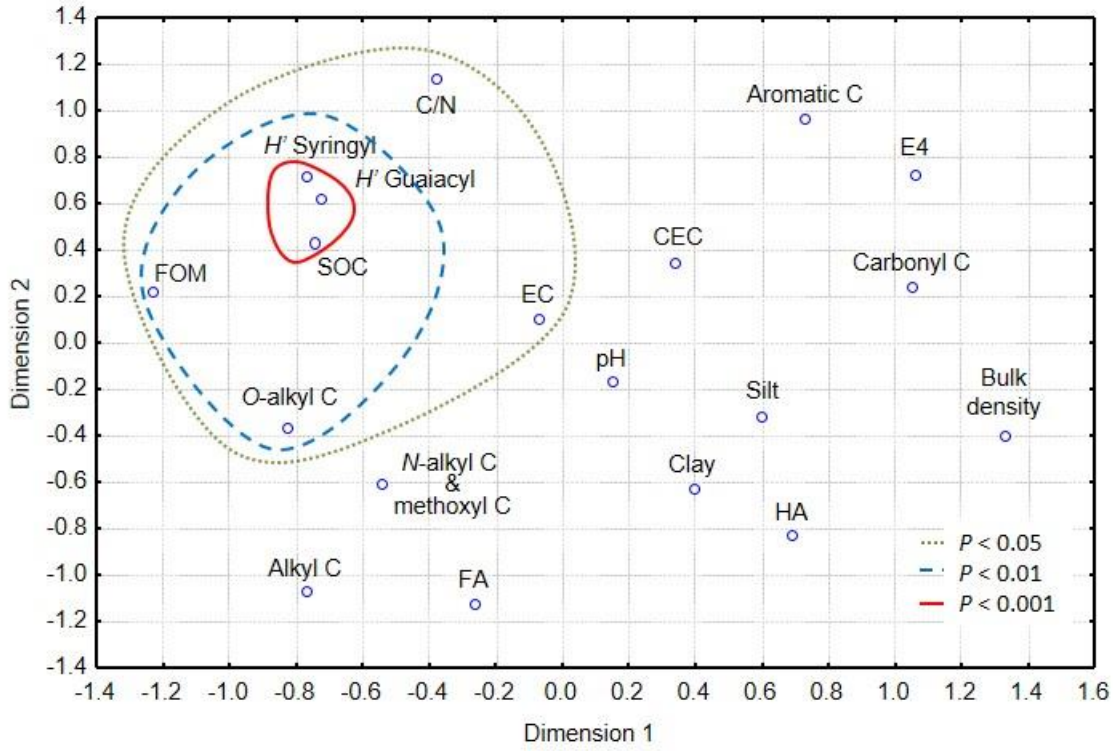


Figure 4.5. Representation, using multidimensional scaling (MDS), of the relationships between different soil properties and the Shannon *H'* diversity index of pyrolytic methoxyphenols. Variables related with progressively less significant *P* values with the soil organic carbon (SOC) are encircled with continuous, dashed and dotted lines, respectively. Variable labels refer to Tables 3, 4, 5, 6 and 7.

In particular, when comparing the correlations between SOC and a large number of variables informing on soil physical and chemical characteristics, or on the concentration of the main soil organic fractions, or even the amounts of the major constituents of HA analyzed by ^{13}C NMR, it is found that the *H'* biodiversity indices of the methoxyphenols are the only variables very close to the SOC, with correlation indices $P < 0.001$ (Fig. 4.5).

At a next level, other variables are only correlated at $P < 0.01$ with the SOC, e.g., FOM and percentage of *O*-alkyl C in HA, i.e., variables indicative of accumulation of litter-derived organic matter. Finally, other soil variables as C/N ratio and EC correlated with SOC only at $P < 0.05$ level.

4.4 Discussion

Recent studies have centered on factors such as physical protection of SOM by soil minerals that affect on the retention of SOM in different depths (Schmidt et al., 2011). Nevertheless, in our opinion, the large variability in the SOC content (Table 3) in our samples probably depends on a plethora of factors, which is presumptively reflected by the complex composition of pyrolytic compounds.

In this sense, most recent literature also tend to consider soil C storage as a scale-dependant process largely influenced by local factors, avoiding ideal models based on the impact of global processes. In fact, carbon storage has been considered to be largely dependent on physicochemical factors such as clay content, nutrient status and water regime and temperature on the soil (Oades, 1988). On the other hand, water loss, soil erosion, and soil mismanagement may induce a great decrease of the SOC (Lal, 2004). Factors associated with vegetative cover were major influences on SOC accumulation (Dillon, 2011). In other cases, the key role of the composition of the soil solution has also been considered crucial in determining the final SOC values (Pan et al., 2013).

Assuming that large number of factors is involved in soil C storage, it is worthy to point out that the results in this study indicated than only a limited subset of the whole pyrolytical assemblage i.e., the 12 major methoxyphenols, provided sufficient insight on the SOM concentration, its quality and some additional information on the botanical origin of the major precursors of the SOM.

Apart from the studies considering C stabilization as a process highly dependent on physical interactions in soil, alternative studies have pointed out that the resistance to biodegradation of the SOM may also differ to large extent in laboratory incubation systems in which organo-mineral interactions were not responsible for the origin of the variability (Hernández et al., 2012). In such a system, multivariate models explained the CO_2 release as resulting from a series of structural variables in the HA, none of which having the crucial role. This supports the idea that, as whole, the structural variability of humic matter is an important factor in SOM resilience. In fact, the above mutual correlations between important soil variables with a presumptive role on biodegradation and humification processes, suggest that the molecular composition of the SOM is

more a useful predictor of the SOC levels than the relative amounts of the more or less decomposed, particulate or colloidal SOM fractions.

The large variability in soil types and processes ought also to be taken into consideration. For instance, previous studies on agricultural soils where the vegetation was the same but the geological substrate differed to large extent (Hernández & Almendros, 2012) evidenced the expected importance of the accumulation of amorphous minerals in justifying low CO_2 release rates and, conversely, high SOM levels in the soils. In particular, the entrapment of SOM on microporous matrices under local anaerobic conditions has been recently postulated as a main process enhancing preservation of SOM of microbial origin in volcanic ash soils (Buurman et al., 2007). Nevertheless, even in this extreme scenario with volcanic minerals with recognized protective role of SOM, much of the total variance in the resulting resilience was explained by structural factors. This effect was more conspicuous when the total aromaticity—with poor forecasting potential as regards SOM levels (Hernández & Almendros, 2012)—was studied after isolating the aromaticity due to lignin constituents (i.e., methoxyphenols) from that due to condensed, alkylbenezenoid structures, including pyrogenic organic matter. In this study, the fact that some topsoils accumulate more organic matter than others is explained by the structure of the SOM, reflected in the biogenic methoxyphenol signature which also coincides with a low degree of alteration of macromolecules in soil (mainly lignins and suberins).

4.5 Conclusions

The analysis of the pyrolytic methoxyphenol signature of continental Mediterranean soils suggests that a relation exists between the diversity calculated from the relative concentration of the 12 major phenols and the total SOM concentration. This finding is interpreted as C storage behaves as a soil emergent property, which is reflected by the sources, complexity and interactions of the SOM constituents at a molecular level. Different statistical analysis (simple regression, PLS, MDS) coincide in pointing out to a relation between C storage in soil and the methoxyphenol complexity. Apart from this, the close relationship between the vegetation type and the soil pyrolytic methoxyphenol signature was corroborated by discriminant analysis. Although

no cause-to-effect relationships are postulated from the fact that the biodiversity in the methoxyphenol assemblages parallel the soil C concentration, it is clear that this fact ought to be considered in the current controversy on the origin of SOM recalcitrance, which in our samples can be expressed as a function of pyrolytic proxies informing on the structural complexity of the SOM.

Capítulo 5:

Chemometric assessment of soil organic matter storage and quality from humic acid infrared spectra

Este capítulo ha dado lugar a la publicación: Jiménez-González M.A., Álvarez A.M., Carral P., Almendros G., 2019. Chemometric assessment of soil organic matter storage and quality from humic acid infrared spectra. Science of the Total Environment 685:1160–1168. https://doi.org/10.1016/j.scitotenv.2019.06.231

Abstract

The knowledge of biogeochemical mechanisms involved in soil organic carbon (SOC) storage is crucial to control its release to the atmosphere. In particular, the chemical composition of soil organic matter (SOM) plays an important role in the performance of the C storage and resilience in soils. The structural information provided by infrared spectroscopy (IR) of soil humic acid (HA) was used in the assessment of the C storage potential of 35 Spanish soils. Partial least squares (PLS) regression using the intensities of the points of the IR spectra of the HAs (4000–400 cm^{-1}) as descriptors shows that a relationship exists between IR spectral pattern and the SOC content. This was also the case for E4 (humification index based on HA optical density at 465 nm). In addition, the chemical characteristics of the HAs correlated with the SOC levels were identified from digital data treatments of the IR spectra. Additional application of principal component analysis (PCA) and multidimensional scaling (MDS) suggested that bands assigned to carboxyl and amide structures were characteristic in HAs from soils with low C content, whereas HA spectra from soils with high C levels showed a conspicuous band pattern suggesting structural units of lignin from slightly transformed plant residues. The spectral profiles were analyzed in detail by an approach based on digital subtraction of IR spectra obtained by averaging those from HAs extracted from soils in the upper and lower quartiles of the SOC distribution. The results showed that significant relationships exist between the molecular composition of HAs and SOC levels and E4 values in a way in which aromatic, carboxyl and amide groups were predominant in HAs from soils with low SOC content, whereas lignin-derived structures were more characteristic of HAs from soils with high SOC content.

5.1 Introduction

Global change presumptively produced by increasing carbon dioxide (CO_2) emission to the atmosphere is a well-publicized subject of research from different scientific fields. In fact, soil represents one of the major C reservoirs at the Earth's level (Batjes, 1996; Lal, 2004), then the study of the dynamics of the soil organic carbon (SOC) and the optimization of soil management practices is interesting in terms of preventing increased emissions of CO_2 from soil (Pizzeghello et al., 2017; Hernández et al., 2019). In fact, even a small variation in the SOC pool may be reflected in the resulting atmospheric CO_2 concentration (Lal, 2004).

Biogeochemical factors responsible for SOC storage are only partially known: the stabilization of soil organic matter (SOM) by organo-mineral complexation, or physical protection including microencapsulation plays an important role in enhancing SOC sequestration (Spaccini et al., 2002; Spaccini & Piccolo, 2012; Simonetti et al., 2017). On the other hand, stabilization mechanisms depending on the chemical composition of the SOM are also very important (Solomon et al., 2007; Jiménez-González et al., 2017, 2018). In fact, SOM chemical composition depends on a plethora of biochemical and abiotic reactions such as the occasional but relevant influence of wildfires (Jiménez-González et al., 2016; Pereira et al., 2016). Most research on SOM has progressed significantly taking advantage of studies on the operationally defined humic substances (Spaccini et al., 2002; Song et al., 2014). Such humic substances can be divided into three fractions: fulvic acids (alkali- and acid-soluble fraction), humic acids (HAs) (alkali-soluble and acid-insoluble) and humin (alkali- and acid- insoluble). In particular the HA fraction, which represent one of the most structurally complex pools of the SOM, has been the subject of extensive research on its composition, reactions and hypothetical structural models (Schnitzer & Khan, 1972; Hayes et al., 1989; Hayes, 1991; Stevenson, 1994; Piccolo, 2002; Piccolo et al., 2018). In fact, recent studies take advantage of the classical idea that research on the molecular structure of humic substances represents a source of still unexplored information about biogeochemical processes (Baveye & Wander, 2019). The transformation of SOM and in particular the HA formation pathways depend on different environmental factors (Parton et al., 1987) and are expressed by means of progressive changes in the macromolecular

structure of the SOM (Stevenson, 1994). In fact, the molecular composition of the HAs differs from that of the precursor plant- or microorganism-derived biomass (Miller et al., 2015; Tadini et al., 2015) and the extent of such structural differences as regards biomass could be postulated as a valid surrogate of the SOM maturity, humification degree, or SOM quality (Almendros et al., 2018; Jiménez-González et al., 2018).

Using non-destructive techniques such as infrared (IR) spectroscopy and solid-state nuclear magnetic resonance (^{13}C NMR) it is possible to perform an exploratory analysis of soil characteristics associated with the specific HA chemical composition. Current studies have used mid-IR spectra to predict soil properties (Madari et al., 2006; Janik et al., 2007; D'Acqui et al., 2010; Cécillon et al., 2012), or to study dissolved organic matter in aquatic ecosystems (Abdulla et al., 2010). Typical bands of IR spectra have been systematically ascribed to different SOM chemical constituents and used as valuable semiquantitative descriptors of the impact of several types of environmental factors on SOM characteristics (MacCarthy & Rice, 1985; Fernández-Getino et al., 2013; Miralles et al., 2015) i.e., 1720, 1620 cm^{-1}, associated to carboxyl and aromatic groups, 2920, 1460 cm^{-1} bands for alkyl structures, 1650, 1640 cm^{-1} bands related with protein, or 1510, 1460, 1270, 1230, 1130, 1030 cm^{-1} coinciding with typical lignin structural units (Fengel & Wegener, 1984; Yang et al., 2011). Using the intensity of the above bands combined with quantitative integration data of the different regions of the ^{13}C NMR pectra of the HAs, it is possible obtain a reliable insight of the variability in the HAs' chemical constituents. Nevertheless, in the case of IR spectroscopy, further refinement based on spectral data treatments such as obtaining derivative spectra, autodeconvolution, or resolution enhancement (Almendros et al., 1992) can be very helpful for quantitative assessments, exploratory spectral pattern recognition or to improve the significance levels of chemometric models to relate spectroscopic data with different soil properties as could be SOC content or SOM quality. Although some of these treatments modify IR peak intensities, this effect is the same for homologous peaks of spectra processed by the same procedure. In particular, second derivative IR spectra are usually chosen due to it allow accurate measurement of peak intensities (Luinge et al., 1995; Viscarra Rossel, 2008).

This research includes the application of partial least squares (PLS) regression to explore soil properties reflected in the mid-IR pattern (4000–400 cm^{-1}) of soil

HAs. The principal goal is to explore if there is a relationship between both the SOC content and E4 index, and the IR spectral signature of HAs isolated from the corresponding soils. In a second stage, this research aims to differentiate any specific spectroscopic pattern of HAs accumulated in soils with high potential for SOC storage, with that of HAs from C depleted soils.

5.2 Materials and Methods

5.2.1 Study area

Thirty-five topsoil samples (0–10 cm) were collected from different Spanish ecosystems developed under widely differing environmental characteristics. The general characteristics of soil sampling areas were reported elsewhere (Jiménez-González et al., 2017, 2018). The soils were selected intending to cover a wide range of variability in SOC content (17–157 g C·kg soil^{-1}) due to explaining the variability in this soil property is a main objective of this work. Three sampling points were selected for each soil area, the samples were collected from the A horizon after removing the litter layer. Finally, composite soil samples were prepared by mixing soil subsamples from three different points, and the resulting materials were air-dried at room temperature for a week and homogenized to < 2 mm.

5.2.2 Soil physical and chemical analysis

The soils were classified according to the World Reference Base for Soil Resources (2014) system. Soil texture was determined by the densimeter method (Bouyoucos, 1927). Soil pH was measured in soil:water suspension (1:2.5, w:w) with a pH 7 pH-meter (XS Instruments; Carpi, MO, Italy). The electrical conductivity (EC) was determined with a COND7 conductivity-meter (XS instruments; Carpi MO, Italy) using soil:water slurry (1:5, w:w). The SOC concentration was determined by wet chemical oxidation with 1*M* potassium dichromate (Nelson & Sommers, 1982; Walkley & Black, 1934), and the nitrogen (N) by micro-Kjeldahl digestion (Prince, 1945). The water holding capacity (WHC) of the soils was measured in the laboratory at atmospheric pressure. Tables 3, 4, 6 and 7 show some properties of the soils and its corresponding HAs.

5.2.3 Isolation and characterization of humic acids

The first step of the SOM fractionation consisted of removing free organic matter (FOM, particulate SOM fraction) by flotation in 2*M* H_3PO_4. The humic extract was extracted with 0.1*M* $Na_4P_2O_7$ and 0.1*M* NaOH in reducing atmosphere. The HAs were separated from fulvic acids (FAs) by precipitating the humic extract with 6*M* HCl. The HAs were purified by high-speed centrifugation (46000 × g) followed by dialysis (Visking tube size 6, Medicell International Ltd.) to remove soluble salts.

The optical density (E4) of the HAs was determined at 465 nm in HA solutions in 0.01*M* NaOH at a concentration of 0.2 mg·cm^{-3} (Kononova, 1982), this index is related with the humification degree and has been considered a reliable surrogate of the aromaticity of HAs (Traina et al., 1990; Tinoco et al., 2015). The main HA structural groups were quantified using solid-state ^{13}C NMR spectroscopy using a Bruker Avance 400 MHz instrument, operating at a frequency of 100.63 MHz with 4 mm O.D. zirconium oxide rotors with Kel-F caps. The cross polarization (CP) was used during magic angle spinning (MAS) of the rotor at 12.5 kHz. Between 5000 and 6000 scans were accumulated with a pulse delay of 300 ms. Tetramethylsilane was used to calibrate the ^{13}C chemical shifts (0 ppm). To circumvent Hartmann-Hahn mismatches, a ramped 1H pulse was applied during the 1 ms contact time. The spectral ranges were described according to the body of classical literature (González-Vila et al., 1983; Knicker, 2011): alkyl + α-amino C (0–45 ppm); *N*-alkyl + OCH_3 C (45–60 ppm); *O*-alkyl C (60–110 ppm); unsubstituted aromatic C (110–140 ppm); heteroaromatic C (140–160); carbonyl C (mainly carboxyl + amide, 160–220 ppm). In this study, special attention was paid to the unsubstituted aromatic C region (110–140 ppm). The ^{13}C intensity distribution was determined by integrating peak areas over the above-mentioned chemical shift regions.

5.2.4 Infrared spectroscopy

Fourier transform-infrared spectra were acquired with an Agilent Cary 630 FTIR spectrophotometer at a wavelength range of 4000–400 cm^{-1} and a resolution of 4 cm^{-1}. Potassium bromide (KBr) pellets containing 2 mg of powdered HA and 200 mg of KBr were scanned. Spectral data were background corrected to a reference spectrum prior to every measurement and some spurious bands, such as those from atmospheric CO_2 were removed. Resolution enhancement

techniques (Almendros et al., 1992) are very useful in case of typical broadband IR spectra such as those from HAs, these spectral data treatments revealed peaks not evident in the original spectra. The method used is based on subtracting from the raw spectrum a positive multiple of its 2nd derivative (Fig. 5.1). Finally, 2nd derivative spectra, which show a systematic pattern of peaks and valleys and a straight virtual baseline, were chosen for semi-quantitative measurement of peak intensities (Fernández-Getino et al., 2013).

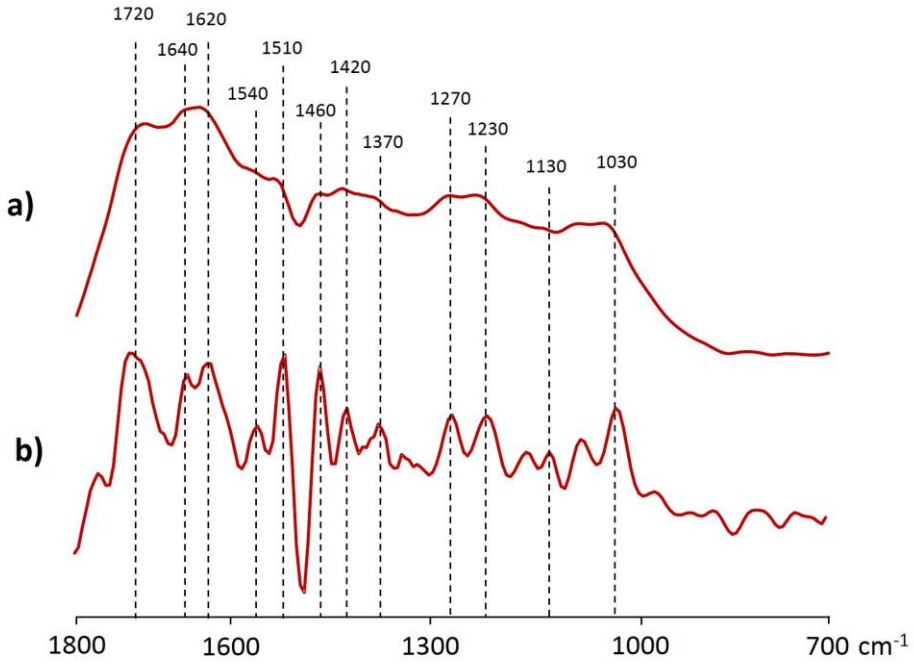


Figure 5.1. Infrared spectrum of humic acid (HA) from sample 34 (a) compared with the corresponding resolution enhanced spectrum (b).

5.2.5 Statistical analysis

Partial least squares regression using the ParLeS software (Viscarra Rossel, 2008) was used to explain the variance of the SOC content (as dependent variable) in terms of a large series of descriptors (independent variables) consisting of the spectral points (4000–400 cm^{-1}) after mean centering. The 2nd derivative transformation of the IR spectra was selected amongst a series of standard preprocessing and pretreatment algorithms included in the program for model optimization. The PLS was applied to explore the possible relationship

between the IR spectrum and SOC and E4 values, the fact that a model is possible shows that a relationship exists between them. The root mean squared error (RMSE) and the Akaike's information criterion (AIC) (Akaike, 1974) were used to select PLS models with the minimum number of factors or latent variables (LVs) i.e., to prevent overfitted spurious models. Then, the results were shown in a cross-validation plot.

In a second stage, and in order to identify spectral features characteristic of the IR spectra of HAs from soils with different SOC levels, the spectra were ordered according the C content of the corresponding soils and two subsets were considered for comparison, i.e., soils above vs. below the median of the C levels, or “soils behaving as C-sinks” vs. “soils with low SOC levels”. Finally, samples were classified by quartiles (Q1 to Q4), after observing that the differences were the same but more important in the case of quartiles. The average of the IR spectra from the HAs of soils in the 1st quartile (Q1, samples with highest SOC content) and the average spectra of IR from the HAs from soils with C levels in the 4th quartile (Q4, samples with lowest SOC content) were calculated. The subtraction of the average spectra from HA samples of these quartiles was carried out to recognize the characteristics spectral bands prevailing in each subset. A similar process was carried out using the E4: the spectra were ordered according the value of E4 of the HAs, and then IR intensity values of average spectra of samples in Q1 and Q4 were subtracted.

In addition, and in an attempt to obtain more perceptual plots illustrating the spectral features of HAs from soils classified according to SOC or E4 values two independent spectral traces, respectively highlighting the positive or negative intensities in the above subtracted spectra were obtained. For this purpose, the average spectrum from HA samples in Q1 is multiplied by a factor suitable to make equal to zero the difference value at the most intense valley in the subtracted spectrum: then the resulting subtracted array (Q1–Q4) exclusively contains positive values, i.e., a virtual spectrum consisting of the characteristic bands prevailing in HA samples from the Q1 set (Manders, 1987; Martínez et al., 1990). The same operation was carried out to obtain the spectrum emphasizing the bands prevailing in HAs from soils with low SOC content, corresponding to the negative subtraction values. In this case the data array in Q1 spectrum was multiplied by a factor to obtain a spectral array where the highest value in the subtracted spectrum would coincide with the zero value,

consequently all peaks in the subtraction are negative. Finally, the resulting array was multiplied by −1 to plot a positive trace. The Student *t* test was calculated for each spectral point to illustrate the significance level ($P < 0.05$) of the spectral intensities of sample sets with different SOC levels.

The relationships between the SOC content and other soil and HA properties were examined by principal component analysis (PCA, correlation matrix, no rotation) using as variables the normalized intensities of the principal IR bands measured in the 2nd derivative and the Statistica ver. 7.1 software. This multivariate data treatment was used aiming to independently check the different contributions of individual IR bands to the variability of SOC and E4. These properties were processed as supplementary variables, i.e., those which are not used to calculate the ordination axes. The principal components were calculated using only the intensity of the selected IR bands as principal variables. Supplementary variables are represented in the space according to the components calculated from the principal variables (Legendre and Legendre, 1998).

Finally, multidimensional scaling (MDS, Kruskal, 1964) was applied to study mutual relationships between the intensity of the main IR spectral bands and soil variables, using the 1-Pearson *r* correlation index as a measure of similarity. Therefore, the proximity of the scores so calculated for the variables in the plane is considered proportional to their mutual correlations.

5.3 Results

5.3.1 ^{13}C NMR spectra

The high variability in the chemical composition of the HAs was suggested by the solid-state ^{13}C NMR spectra. The 0–45 ppm range, associated to alkyl C, was a major region of the ^{13}C NMR spectra, accounting for 25.5 to 41.5%. The signal intensity corresponding to the 45–60 ppm region (*N*-alkyl + OCH_3) showed low variation (7.9 to 11.8%). The lowest intensity in the 60–110 ppm region (*O*-alkyl C) was 14.3% and the highest value reached 27.4%. The 110–140 ppm region, corresponding to unsubstituted aromatic carbons showed values from 11.9 to 27.6% whereas the 140–160 ppm region for heteroaromatic

carbons ranged between 4.3 and 7.4%. Finally, the carbonyl region, between 160 and 220 ppm showed values ranging from 9.4 to 14.8%.

5.3.2 Partial least squares regression

Figure 5.2 shows the leave-one-out cross-validation plots for the PLS prediction model obtained for SOC (i.e., SOM quantity) and E4 (i.e., SOM quality). In the best model selected, the values of RMSE and AIC suggested including up to 6 LVs in the model for SOC prediction and 2 LVs for E4. In the cross-validation plot for the PLS model using the SOC values there was a significant correlation between predicted and observed values, with R^2 = 0.626. In the case of E4 values, a higher correlation (R^2 = 0.886) than for SOC was obtained between the experimental values and the values predicted by the model.

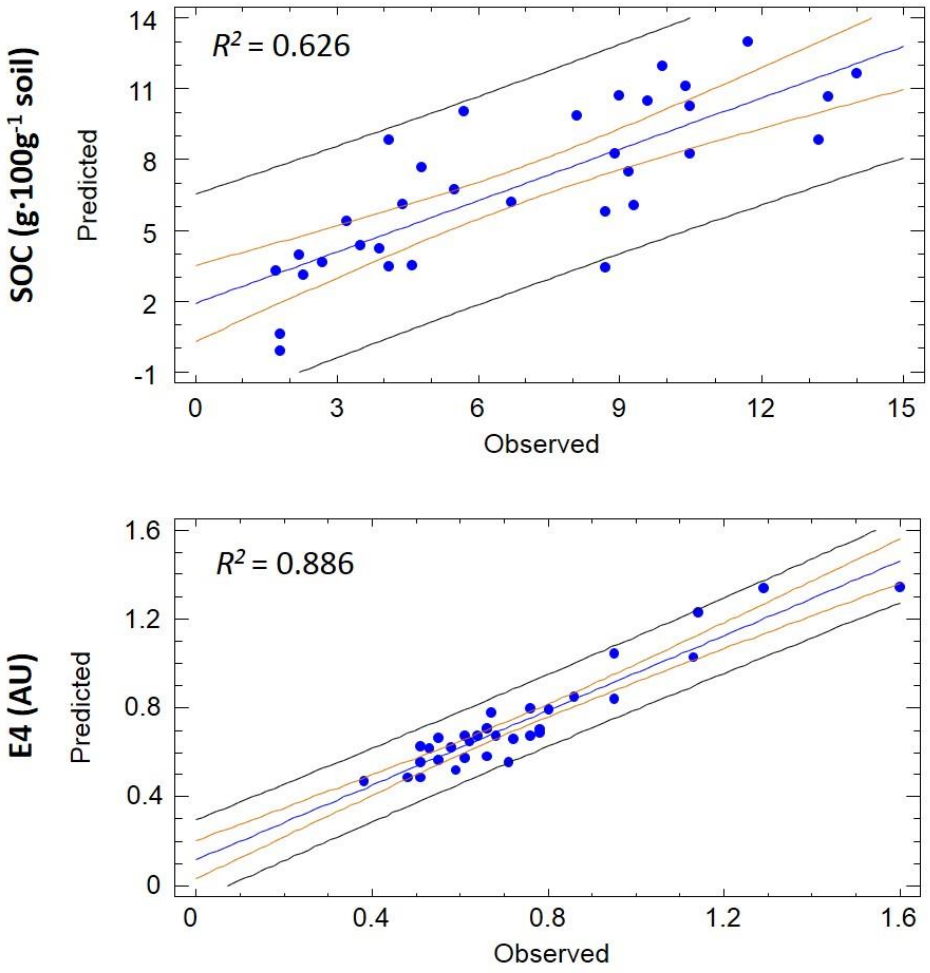


Figure 5.2. Observed versus predicted values of soil organic carbon (SOC) and optical density at 465 nm of soil humic acids (E4) from Mediterranean soils calculated by partial least squares regression (PLS). Inner (orange) and outer (black) lines indicate 95% confidence and prediction limits, respectively.

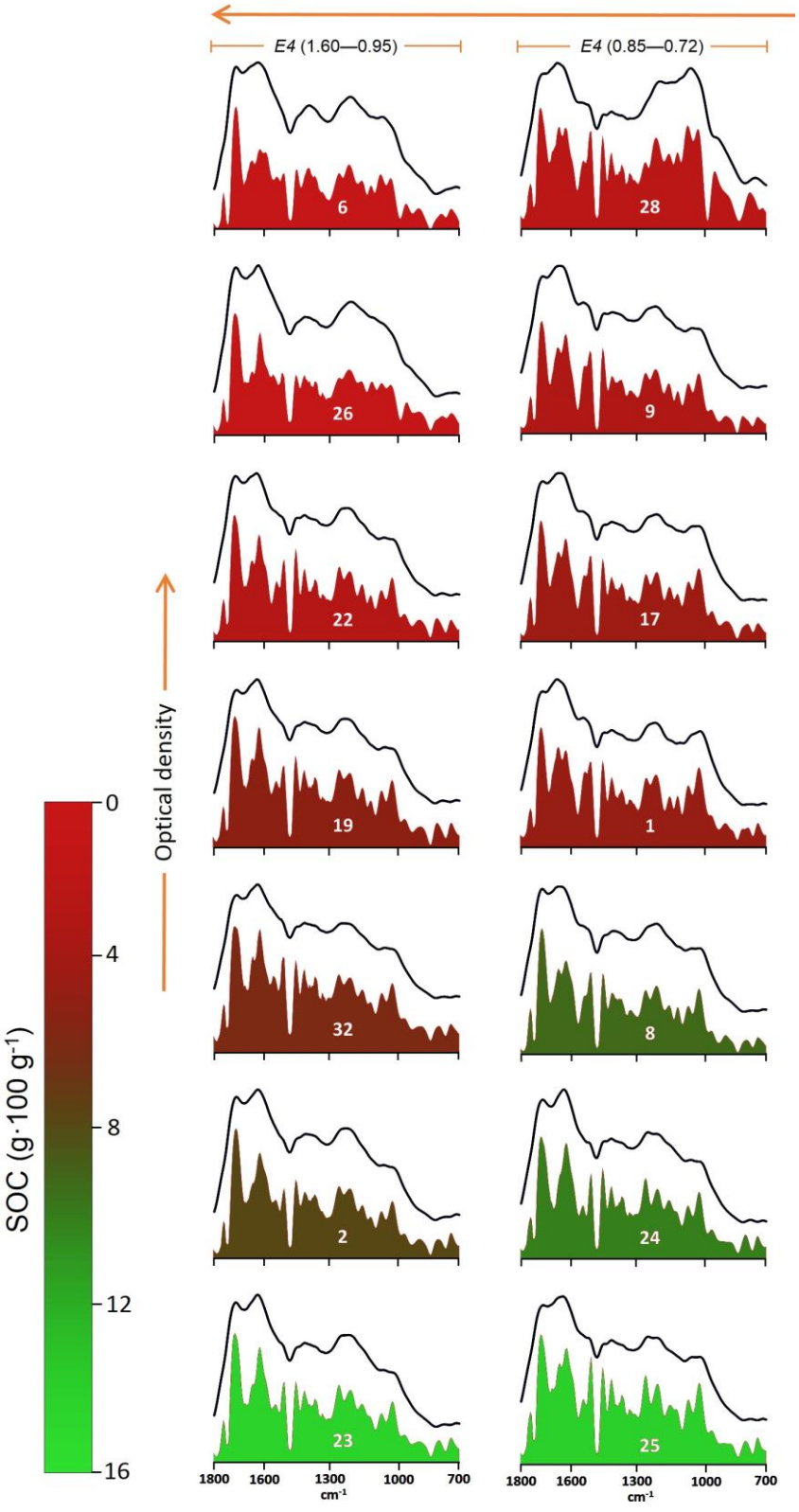
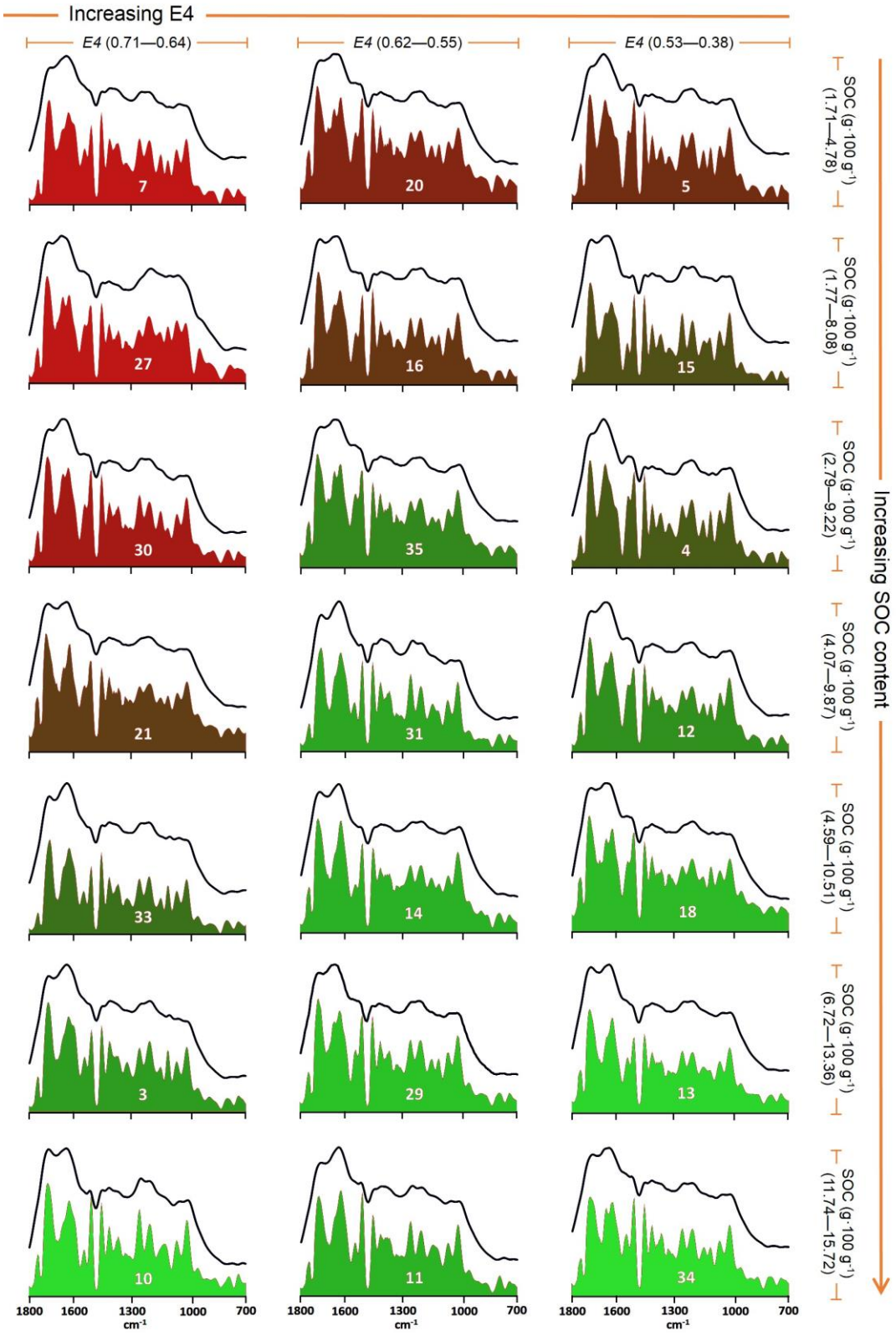


Figure 5.3. Representation of infrared spectra and resolution enhanced spectra of humic acids from Mediterranean soils. Samples are ordered horizontally in function of E4 value and vertically in function of increasing content of soil organic carbon (SOC). The samples are colored according the SOC content. Numbers of the spectra refer to Table 1.


Increasing E4
E4 (0.71—0.64)
E4 (0.62—0.55)
E4 (0.53—0.38)
7
20
5
27
16
15
30
35
4
21
31
12
33
14
18
3
29
13
10
11
34
1800 1600 1300 1000 700
cm-1
SOC (g·100 g-1) (1.71—4.78)
SOC (g·100 g-1) (1.77—8.08)
SOC (g·100 g-1) (2.79—9.22)
SOC (g·100 g-1) (4.07—9.87)
SOC (g·100 g-1) (4.59—10.51)
SOC (g·100 g-1) (6.72—13.36)
SOC (g·100 g-1) (11.74—15.72)
Increasing SOC content

5.3.3 Infrared spectra

The Fig. 5.3 shows the 35 spectra superimposed to the corresponding resolution enhanced IR spectra. These spectra are ordered according E4 values and then according the SOC content. It is observed how the spectra with marked lignin pattern are located in the area of high SOC and low E4, whereas HAs from soils with low C levels tend to show strong carboxyl and aromatic bands.

In Fig. 5.4a, a comprehensive representation of the differences between spectroscopic patterns of HAs from soils with high levels of SOC as regards those with low potential for SOC storage, i.e., exposed to more severe risks of desertification, was obtained by subtracting the two spectra obtained after averaging spectra corresponding to sample sets with high and low SOC, respectively. This representation of the differential spectral features of HAs from soils in Q1 vs Q4 shows positive and negative peaks: the positive values correspond to bands predominant in the IR spectra of HAs isolated from soils with high SOC content (Q1). In the latter case, it is observed the bands at 1510, 1460, 1420, 1270, 1130 and 1030 cm^{-1}. On the other side (Q4), other bands are predominant In HAs from soils with low SOC levels, i.e., bands at 1540 cm^{-1} which are related with amide groups, or 1720 cm^{-1} bands linked to C=O groups.

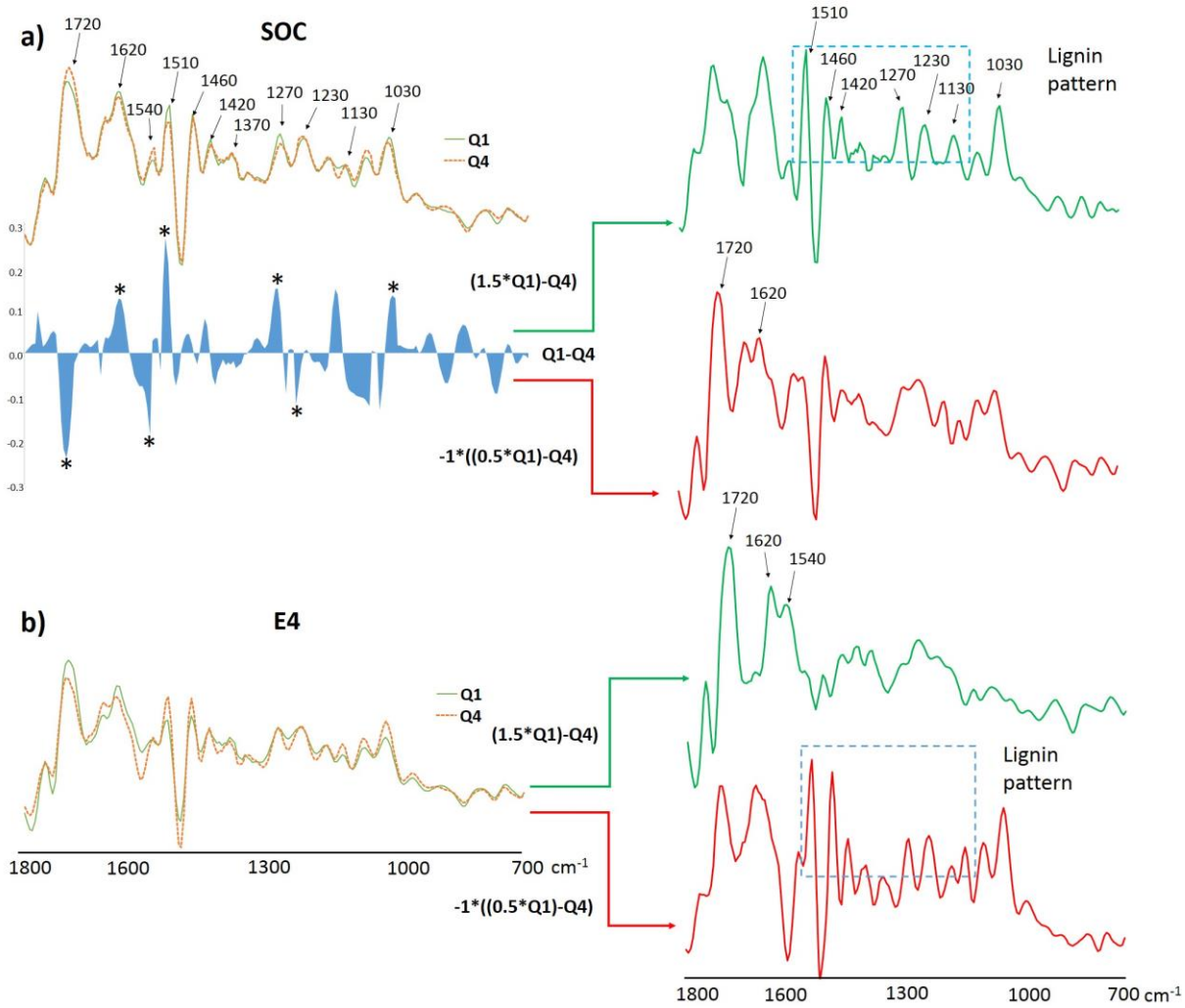


Figure 5.4. Comparison of IR patterns of humic acids (HA) in terms of a) soil organic carbon (SOC) levels and b) optical density at 465 nm of soil HA (E4). The subtraction values of spectral intensities between average IR spectra from HAs in extreme quartiles of the SOC distribution in the corresponding soils is represented in a). The Q1–Q4 positive values correspond to bands predominating in C-rich soils (green trace), whereas negative values correspond to bands predominating in C-depleted soils (red trace). The asterisks (*) show values statistically different ($P < 0.05$) between band intensities of the IR spectrum in the soils with comparatively high C storage potential as regards soils comparatively poor in C (Q1 and Q4) according to the Student's *t* test. For a more perceptual spectroscopic interpretation of the organic constituents prevailing in each of the HAs from Q1 and Q4 soil sets, two independent spectra were calculated after zero-correction baseline by multiplying by a linear factor, showing the different composition of the corresponding HAs from samples of high (green trace) and low (red trace) values of the dependent variables SOC and E4, respectively.

In the representation of the spectral profiles in which the positive and negative values are respectively extracted after zero-correction baseline by multiplying by a linear factor, the different composition of the corresponding HAs (from C-rich and C-poor soils) became much more evident. The first spectrum representing the positive region, showed the well-defined lignin pattern with sharp peaks (Fig. 5.4a). The second spectrum, which emphasizes the negative region of the subtraction spectrum, showed a prominent peak at 1720 cm^{-1} (C=O groups) and a broad band at 1620–1640 cm^{-1} (aromatic, unsaturated, quinone and amide structures). The same conclusion, but with opposite signs was obtained in the subtraction spectra obtained using the E4 ranks as classification factor (Fig. 5.4b). Negative values which are predominant in HAs with low E4 values define a spectrum with a marked lignin pattern whereas the spectrum displaying the positive values after the subtraction showed marked bands at 1720, 1620 and 1540 cm^{-1}.

5.3.4 Principal component analysis

The PCA results are shown as a biplot in Fig. 5.5. The two first components obtained using as variables the intensities of the principal bands of the IR spectra explained 27.4% (first component) and 19.5% (second component) of the total variance. Observing the position of the eigenvectors corresponding to the variables (band intensities of the IR spectra) used to calculate the components, it is possible to describe relationships existing between them according the location in the plane defined by the two components selected. The labels of the soil samples were located in the plane according the factors calculated. It is observed that eigenvectors for a set of bands (1510, 1460, 1420, 1370, 1270, 1130, 1030 cm^{-1}) are oriented towards a common region of the factorial space. In the same area of these bands in the plot, other soil properties like SOC, WHC, C/N ratio are located closer. On the other hand, high loading factors for bands at 1720 and 1620 cm^{-1} are defined in the opposite region near to properties such as E4, and aromatic (110–140 ppm) and carbonyl C regions of the NMR spectra (160–220 ppm).

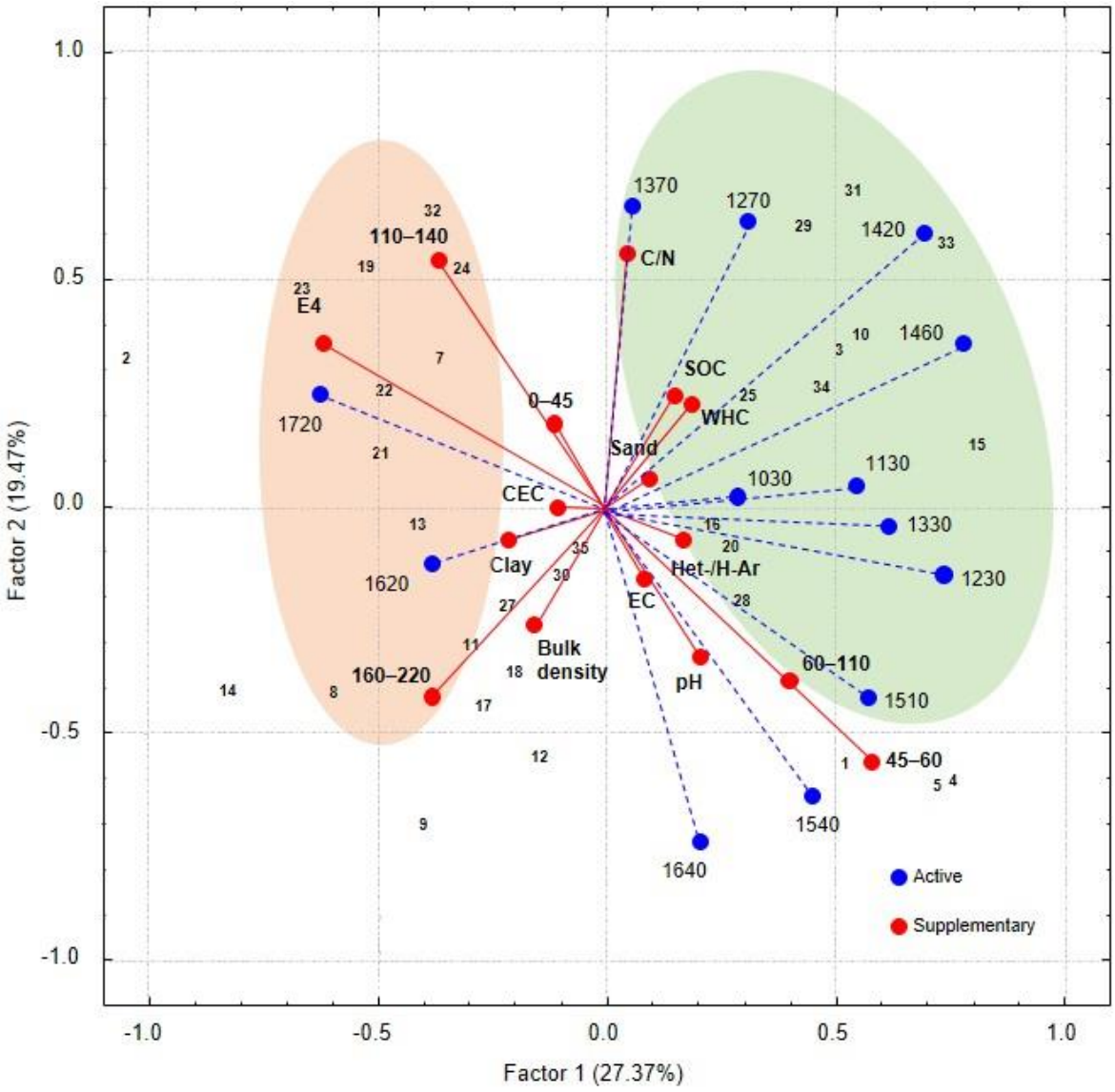


Figure 5.5. Biplot obtained by principal component analysis (PCA), with numbers (1–35) representing soil samples and eigenvectors pointing to the direction of increasing values of the independent variables (infrared (IR) bands, represented with blue dashed lines). A series of soil properties (also considered as dependent variables and processed as supplementary variables) are displayed with solid red lines. The biplot suggest two main clusters with variables correlated with the soil organic carbon (SOC) content (encircled in green) or the aromaticity of the humic acid (HA) (encircled in red), respectively. Variable labels refer to Tables 3, 4, 6, 7. Numbers refer to peak wavenumbers in the IR spectra (1720, 1640, 1620, 1540, 1510, 1460, 1420, 1370, 1330, 1270, 1230, 1130 and 1030 cm^{-1}) and signal area in selected ranges of the ^{13}C NMR spectra (0–45, 45–60, 60–110, 110–140 and 160–220 ppm).

5.3.5 Multidimensional scaling

The scatterdiagram in Fig. 5.6 shows the ordination of the variables using MDS. The final stress level in this analysis was 0.225 (Kruskal, 1964). In general, MDS analysis confirms the mutual relationships between variables suggested by PCA. In this case, the scores for variables related with SOM quality as E4, aromaticity (110–140 ppm) determined by NMR and intensity of the IR band at 1720 cm^{-1} are located closer with a significant correlation of 95% and in the opposite site to SOC level in the scatterplot. The SOC is located close to IR bands 1510, 1270, 1030 cm^{-1} and variables like WHC and Het-/H-Ar with a correlation of 95%. In addition, a set of bands (1460, 1420, 1370, 1130 cm^{-1}) and C/N are also clustering close to SOC in a correlation of 90%.

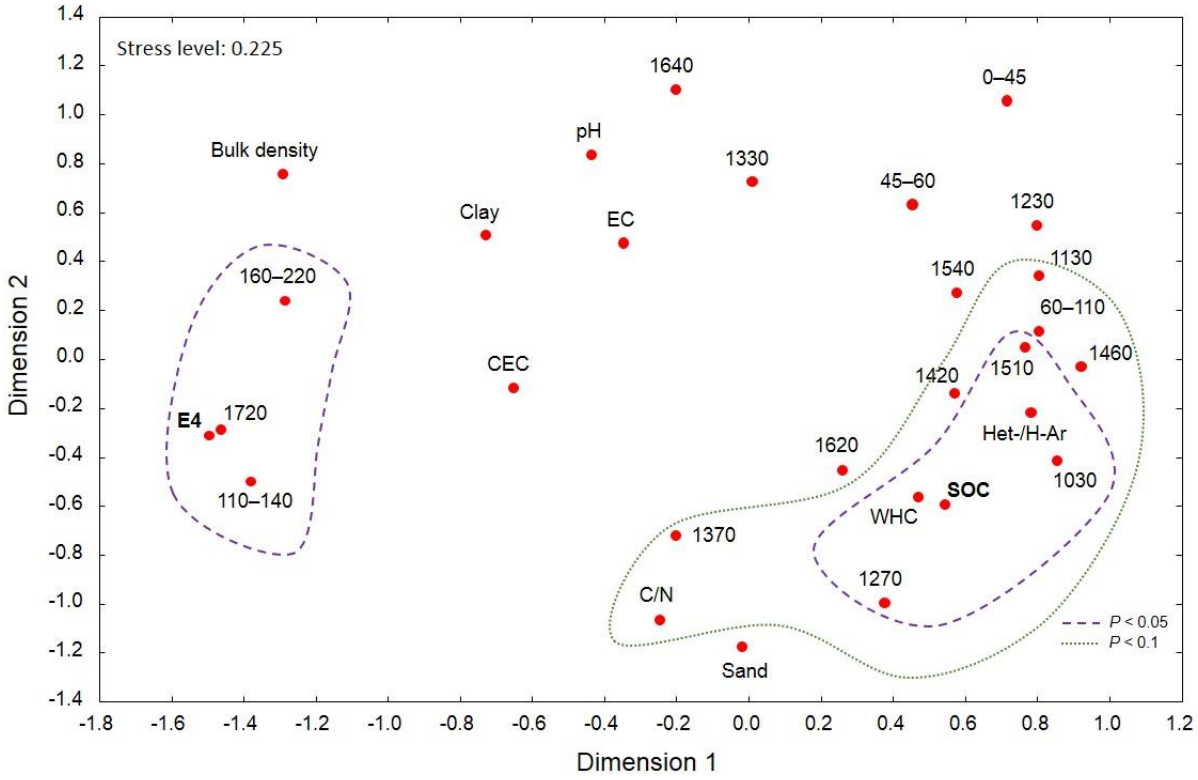


Figure 5.6. Representation, using multidimensional scaling (MDS) of the relationships between different soil properties and quality descriptors of soil organic matter (SOM), and the intensities of the main bands of the infrared (IR) spectra of the corresponding soil humic acids (HAs). Concentric lines indicate the variables positively correlated either with the soil organic carbon (SOC) content or with the E4 at $P < 0.05$ (dashed line) and at $P < 0.1$ (dotted line). Variable labels refer to Tables 3, 4, 6, 7. Numbers refer to peak intensities of the IR spectra (1720, 1640, 1620, 1540, 1510, 1460, 1420, 1370, 1330, 1270, 1230, 1130 and 1030 cm^{-1}) and signal area in selected ranges of the ^{13}C NMR spectra (0–45, 45–60, 60–110, 110–140 and 160–220 ppm).

5.4 Discussion

The analysis by ^{13}C NMR spectroscopy showed clear differences in the chemical composition of the HAs. In particular, some ^{13}C NMR regions such as that for unsubstituted aromatic C (110–140 ppm) showed large variability. The intensity of this region is often considered to parallel the degree of humification (Tinoco et al., 2015). This is due to the fact that constituents of plant and microbial biomass have a comparatively aliphatic character: both the microbial biomass and even lignin from vascular plants, have not such a high proportion of unsubstituted aromatic carbons. This result suggests great differences in the advanced stages in the humification process and could be useful to monitor the resulting maturity of the HAs.

The conspicuous prediction model obtained for SOC concentration exclusively using IR spectroscopic information indicated that a relationship exists between the functional groups of the HAs reflected in the IR spectra and the SOC level in the soils studied. Similar situation was found for E4, which was also predicted by a model exclusively using the IR spectral information. These findings suggest that the IR spectra of the HAs in the soils studied contain useful biogeochemical information about some soil characteristics such as SOC content, but also on qualitative HA properties on the transformation degree of the SOM, such as the E4, which is correlated with the proportions of carboxyl and aromatic structures (Fernández-Getino et al., 2013; Miralles et al., 2015).

The arrangement of the IR spectra in Fig. 5.3 suggested the accumulation of lignin-derived structures in the composition of HAs from soils with the higher SOC content and with low E4 values, whereas in soils with low SOC and high E4 this pattern is not observed. The subtraction (Fig. 5.4) between average IR spectra from soils belonging to quartiles in opposite levels of the SOC distribution (Q1–Q4) illustrates the predominant bands in the IR spectra according to the C storage potential of soils; the positive values (predominant bands in Q1) in the subtracted spectra reveal a characteristic IR lignin pattern viz., 1510, 1460, 1420, 1270, 1130 and 1030 cm^{-1} (Fengel & Wegener, 1984; Miralles et al., 2015). This would indicate that the soils under study with high SOC levels, display IR spectroscopic features of SOM in early transformation stages. On the other hand, negative values (predominant bands in Q4) of the subtracted spectra suggest a higher intensity of bands related with carboxyl

groups (1720 cm^{-1}) and amides (1540 cm^{-1}). It is considered that these structures are characteristic of advanced humification or transformation stages (Fig. 5.7) (Stevenson, 1994; Jiménez-González et al., 2017, 2018; Tinoco et al., 2018). In fact, oxidized SOM with high aromaticity and N content would indicate extensive microbial reworking i.e., classically associated to an enhanced resilience of the resulting SOM (Requena et al., 1996; Almendros & Dorado, 1999; Pendall & King, 2007) driven by biogeochemical processes leading to progressively increased concentration of aromatic and carboxyl groups. This is better reflected in the two spectra illustrating—after multiplying by zero-baseline correction factors—the positive and negative peaks of the of Q1–Q4 spectral subtractions (Fig. 5.4). In the case of the E4 values, we found the opposite situation. Soils with high E4 not only display weak lignin pattern, but also intense 1720 cm^{-1} band, as well as prominent aromatic and amide regions (1620, 1540 cm^{-1} respectively), whereas soils with HAs with low E4 displayed well-defined IR peak lignin pattern. These facts suggest that SOM quantity and quality are linked to a different molecular composition. The HAs from soils with high SOC content tend to accumulate lignin and aliphatic structures suggesting the continuous input of fresh organic matter from vascular plants.

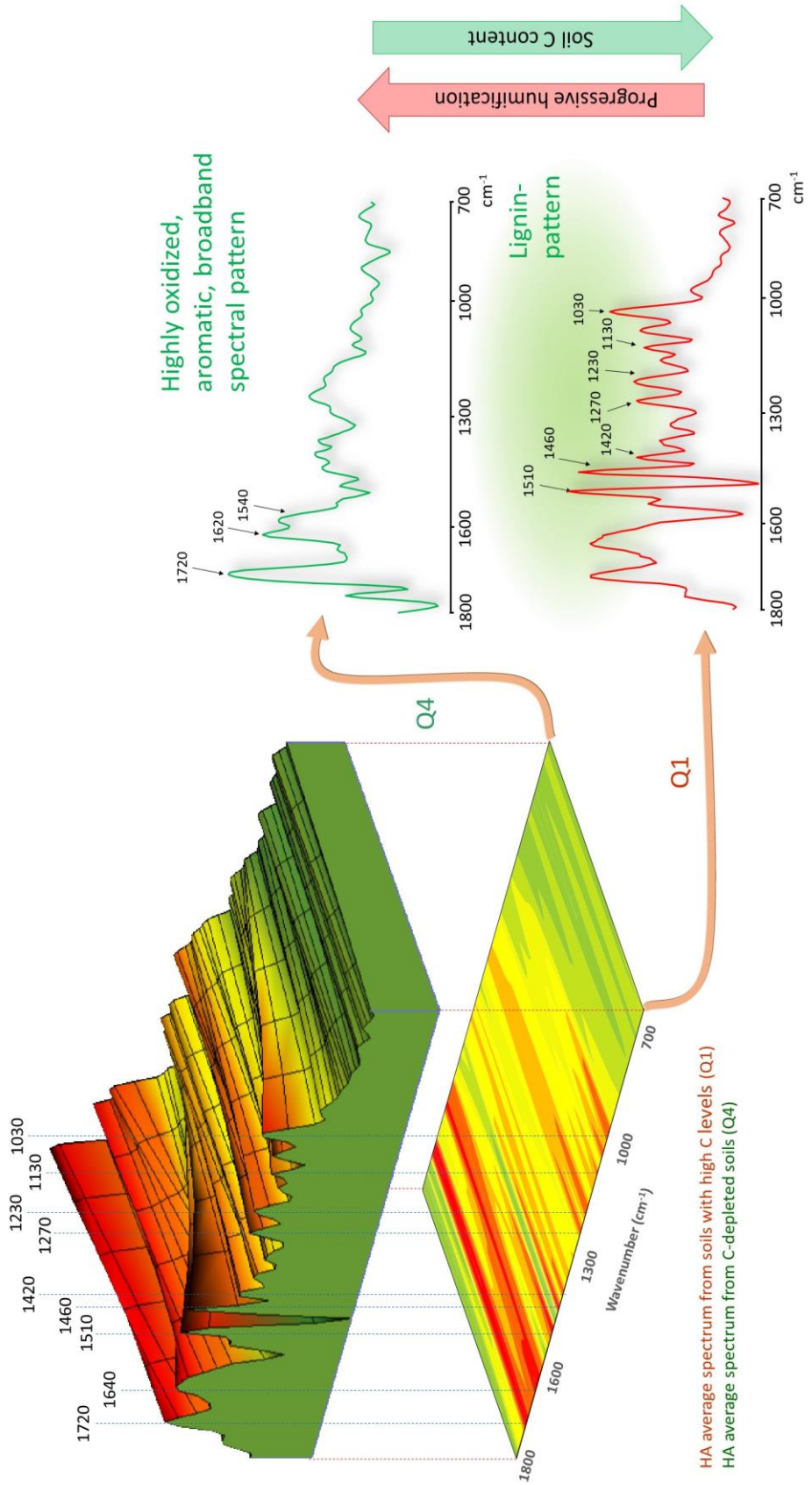


Figura 5.7. Graphical abstract showing changes in mid-infrared spectral patterns of soil humic acids (HAs) in terms of the progressive soil C levels and evolution stages of the soil organic matter.

The biplot showing the arrangement of the studied samples and variables processed by PCA shows two groups of eigenvectors projected onto separate regions of the factorial space and presumably suggesting different C storage processes. Thus, a group of eigenvectors points to the variable corresponding to the SOC concentration, whereas others point to the main processed SOM quality descriptor, E4. In addition, eigenvectors for bands corresponding to lignin structural units (1510, 1460, 1420, 1270, 1230, 1130 cm^{-1}) are grouped in the factorial space, as could correspond to the fact that these bands conform a conspicuous pattern in the IR spectra. The intensity of these bands, which are indicating comparatively slightly transformed SOM (Miralles et al., 2015), are correlated positively with the SOC levels. On the other hand, the projections of the eigenvectors associated to WHC and SOC were very similar, probably due to high SOC levels are connected to improved soil structure and water retention (Nath, 2014; Libohova et al., 2018). The eigenvectors for E4, the intensity of the 110–140 ppm and 160–220 ppm of NMR region were oriented to the opposite side of the plot, coinciding with the scores for the samples in which the SOM evolution presumably reaches more advanced stages. In addition, the close relationship observed between these three variables and the intensities of the 1720 and 1620 cm^{-1} IR bands indicate a high functionality expected from large concentration of aromatic and O-containing functional groups in HAs of high degree of maturity. This coincides with the fact that the eigenvector for E4, a classical index of the degree of transformation of HAs, points to similar zones than those for the unsubstituted aromatic and carbonyl region of NMR, the 1720 cm^{-1} carboxyl and the 1620 cm^{-1} aromatic IR bands. Such facts have been interpreted as the aromatic structures and reactive O-containing groups represent the final products of the humification process, represented by stable structures resistant to biodegradation (Stevenson, 1994). In consequence, the contrasted IR spectroscopic features recognized in the resolution enhanced spectra suggest inverse relationship between quality and quantity of SOC in the studied soils, i.e., soils with low SOC content would contain HAs at comparatively advanced stages of transformation, whereas soils with high SOC levels would contain HAs comparatively richer in slightly transformed and labile structures. This fact also suggests that the routine treatments used for the extraction and purification of HAs are adequate to preserve relevant chemical

information that can be recognized by IR spectroscopy and subsequently related to the origin and formation conditions of the SOM.

Finally, the MDS was useful to analyze the information about HA characteristics reflected in the IR spectra, viz., the close relationship between E4, the intensity of the 1720 cm^{-1} band and the areas of the aromatic (110–140 ppm) and carbonyl (160–220 ppm) ^{13}C NMR regions, and the negative correlation of these variables with the SOC content that was located in the opposite side of the plane (Fig. 5.6). Such pattern of variables supports the suggestion that high SOC levels parallel accumulation of HAs with comparatively labile and slightly transformed biomass structures.

5.5 Conclusions

The analysis by IR spectroscopy of HAs from a set of Mediterranean soils formed under contrasting environmental factors suggests that the HAs from soils with different C levels also have a different molecular composition. The methodological approach suggested, based on previous validation of PLS models to explain the soil C content using the IR spectra of HA as a source of descriptors, followed by digital subtraction between average spectra of HAs from soils with extreme values of C, points out that the SOM levels paralleled the accumulation of HAs consisting of lignin-derived structures, whereas HAs in soils with low C levels present a composition in which aromatic, carbonyl and amide groups prevail. Concerning SOM quality or maturity (using E4 optical density of the HAs as comprehensive indicator) the results pointed to the opposite trend, i.e., the progressive depletion of lignin structures and accumulation of oxidized, aromatic, N-containing structures. In any case, the resolution enhanced IR spectra behaved a reliable source of additional structural information on the impact of environmental processes involved in soil C storage which are reflected in the composition of the corresponding HA fraction.

Capítulo 6:

Ultra-high-resolution mass spectrometric analysis of the molecular composition of humic acids in relation with soil carbon storage

Este capítulo ha dado lugar a un artículo que se encuentra en proceso de revisión: Jiménez-González M.A., Almendros G., Waggoner D.C., Álvarez A.M., Hatcher P.G. Ultra-high-resolution mass spectrometric analysis of the molecular composition of humic acids in relation with soil carbon storage.

Abstract

Long-term stabilization of soil organic matter (SOM) as recalcitrant forms plays an important role in the Earth's carbon cycle. Hence, understanding biogeochemical mechanisms of soil C sequestration is crucial to control its release to the atmosphere. This study aims at investigating the biogeochemical mechanisms of soil C sequestration. An exploratory assessment was carried out on the information on the C storage potential of soils provided by the molecular composition of humic acids (HAs) analyzed by electrospray ionization (ESI) Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS). The application of partial least squares (PLS) regression to the data from FTICR-MS suggested a relationship between the molecular composition of organic matter and the C levels in the corresponding soils. Significant PLS forecasting model for total soil C is obtained using as descriptors the 131 compounds in common in all the HAs detected by FTICR-MS, and its variable importance for projection (VIP) was plotted in the space defined by their atomic ratios using van Krevelen diagrams. The results indicated that significant relationships exist between the HAs molecular composition and the soil organic C levels. The VIP values for the different groups of compounds illustrates how HA contains information about the amounts of C stored in the soil: the HAs in the soils with high levels of organic C have significantly ($P < 0.1$) higher proportions of unsaturated lipid and lignin-derived compounds; on the other side, low soil organic C levels are associated to HAs with comparatively high proportions of saturated lipid compounds.

6.1 Introduction

Soil is the largest carbon reservoir on Earth (Batjes, 1996; Lal, 2004); for this reason, small variations in soil organic C (SOC) content may have a dramatic effect on the atmospheric CO_2 concentration. In fact, the progressive land desertification of soils in Mediterranean areas associated to increasing emission of CO_2 to the atmosphere due to the mineralization of the soil organic matter (SOM) has encouraged recent research attempting to interpret the changes in molecular composition of SOM (Solomon et al., 2007; Faria et al., 2015). This molecular composition of SOM in particular its humic acid (HA) fraction can be a source of information to be related with the humification mechanisms involved in SOM recalcitrance, resulting in the soil potential for C storage (Jiménez-González et al., 2017, 2018, 2019). In consequence, humic substances are among the most investigated topics of soil science (Spaccini et al., 2002; Song et al., 2014). In particular HA, which is one of the most structurally complex SOM pools at least from the viewpoint that its molecular composition to large extent differs from that in the precursor plant or microorganism-derived biomass constituents (Miller et al., 2015; Panettieri et al., 2014; Tadini et al., 2015). According its solubility, humic substances can be operationally divided into three fractions: fulvic acids (alkali- and acid-soluble fraction), HAs (alkali-soluble and acid-insoluble fraction) and humin (alkali- and acid-insoluble fraction) (Stevenson, 1994). Some recent studies have suggested a clear relationship (irrespective to any causal relationship) between the HAs' molecular components, SOC levels and resistance to biodegradation of the SOM (Almendros et al., 2018). The humification processes and transformation of SOM depend on different chemical, physical and environmental factors (vegetation type, climate, pH, geological substrate, etc) (Parton et al., 1987) and this transformation is expressed by means of progressive changes in the molecular composition of the SOM (Stevenson, 1994). In consequence, the origin of SOM from plant or microbial biomass, as well as the extent of its biogeochemical transformation could be scrutinized from the analysis of molecular proxies such as a series of signature compounds that are part of the structure of the HAs. For instance, guaiacyl and syringyl methoxyphenols for lignins from different origin, levoglucosan from pyrolysis of carbohydrates, aliphatic hydroxyacids and diacids from plant polyesters e.g. cutins, or even

cyclic lipid compounds such as diterpenoids, for gymnosperms, triterpenes for most angiosperms, etc (Derenne & Quénéa, 2015).

The organo-mineral interactions and the different protection models of SOM have also been studied as important factors in SOM resilience (Spaccini et al., 2002; Spaccini & Piccolo, 2012; Simonetti et al., 2017). The chemical composition of the HAs may be primarily or secondarily linked to the recalcitrance of this SOM fraction to the extent to which compositional differences in the molecular structure of the HAs are significantly reflected in the levels of organic C content in the soil (Almendros et al., 2018).

A classical problem in the study of the molecular composition of the SOM fractions is the unavoidable structural alteration produced during its pretreatment to be analysed, e.g., wet chemical degradation methods, extraction of SOM by strong acid and bases. In general, destructive techniques such as analytical pyrolysis and wet chemical degradation provided useful information about the nature of the SOM composition. Nevertheless, in some cases the quantitative yields are very doubtful, and these techniques are mainly used for “fingerprinting” humic substances from different origins. In any case, accurate structural analysis cannot be carried out by analysis of the whole unfractionated SOM using classical approaches e.g., infrared spectroscopy (IR) or nuclear magnetic resonance spectroscopy (NMR), that in general do not provide structural information at the level of the units of the macromolecules. The application of “hybrid” techniques such as Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) has shown a considerable potential for fast and direct analysis of molecular compounds in more detailed level (Sleighter and Hatcher, 2007). Although the fragments released by FTICR-MS may not be representative of the total components of the SOM, it is also true that they provide fine molecular-level information that varies significantly depending on the origin of the sample under study. This is especially helpful both in fingerprinting studies as well as for statistical approaches based on information provided by a limited number of compounds with value as source-specific tracers (Jiménez-Morillo et al., 2018). This technique has been recently applied to the compositional research of humic substances (Ikeya et al., 2015; DiDonato et al., 2016) and even on reactions in studies on black carbon and alicyclic aliphatic compounds formation from lignin precursors (Waggoner et al., 2015). In particular, this approach has the advantage that can identify thousands

of compounds with a detailed level of information about its empirical formulas. At first sight, this massive information may represent a limitation in the interpretation of the results, however using the van Krevelen graphical-statistical method (van Krevelen, 1950) to display the molecules in the space defined by its H/C and O/C ratios, it is possible to compare the differentiating features of the SOM molecular structure even in whole soil samples. This graphical representation of the compounds in the H/C and O/C plane is helpful to obtain an insight on their origin, composition and chemical nature (van Krevelen, 1950; Kim et al., 2003; Kramer et al., 2004; Ikeya et al., 2015). After the study by Kim et al. (2003), using this diagram to display compounds detected by FTICR-MS from natural organic matter, a series of authors used this approach to compare different fractions of humic substances from contrasting environments (McKee & Hatcher, 2015; DiDonato et al., 2016; Kamjunke et al., 2017). For exploratory purposes, a series of regions of the van Krevelen diagram have been considered by several authors as corresponding to defined groups of organic compounds with characteristic H/C and O/C ratios (aromatic, condensed aromatic, lignin-derived, carbohydrate, etc) (Kramer et al., 2004; Ikeya et al., 2015). This combination of the FTICR-MS and the van Krevelen diagram has provided a very perceptual approach to study the natural transformation and evolution paths of SOM and fossil organic resources (Ikeya et al., 2015; Kim et al., 2003).

This research is based on the application of FTICR-MS to HAs from contrasting soils with a large variability in ecological, physical and chemical soil properties and specifically SOC content. The principal goal is to explore possible structural relationships between HA composition and the SOC storage. In addition, and as high C content in soil could in some cases be associated to a particular resistance to the biodegradation of the corresponding SOM, this study also aims to describe which molecular constituents behave as SOM resilience descriptors. Moreover, the data obtained would be useful to infer systematic structural relationships illustrating the extent to which HAs retain environmental information on soil properties reflected in the resulting SOC concentration.

6.2 Materials and Methods

6.2.1 Sampling sites

Topsoil samples (0–10 cm) were collected from 35 different ecosystems mainly from continental Spain with a large variability in climatic conditions, vegetation and geological substrate (Tables 1 and 2). The soils were selected intending to cover a wide range of variability in SOC content. The soils were classified according to the IUSS Working Group WRB (2014). Three sampling points were selected for each soil sample, the soil samples were collected from the A horizon after removing the litter layer in each point. In a second stage, a composite soil sample was prepared by mixing soil material from the three sampling points, then the resulting material was air-dried and homogenized to < 2 mm.

6.2.2 Laboratory analysis

The free organic matter (FOM, particulate soil organic fraction) was removed by flotation in 2*M* H_3PO_4. The extractable humic substances were isolated with standard method using 0.1*M* $Na_4P_2O_7$ and 0.1*M* NaOH in reducing atmosphere. The HAs were separated from the fulvic acids (FAs) by precipitating the total humic extract with 6*M* HCl. The HAs were purified by high-speed centrifugation (46000 × g) followed by dialysis (dialysis tubing-Visking size 6, Medicell International Ltd.) to remove the salts. The main HA structural groups were studied using solid-state ^{13}C NMR spectroscopy. The spectra were obtained in a Bruker Avance 400 MHz, operating at 100.63 MHz with zirconium oxide rotors of 4 mm O.D. with Kel-F caps. The cross polarization (CP) was used during magic-angle spinning (MAS) of the rotor at 12.5 kHz. Between 5000 and 6000 scans were accumulated with a pulse delay of 300 ms. Tetramethylsilane was used to calibrate the ^{13}C chemical shifts (0 ppm). To circumvent Hartmann-Hahn mismatches, a ramped 1H-pulse was applied during the 1 ms contact time. The spectral ranges were chosen according to the body of classical literature (González-Vila et al., 1983; Knicker, 2011; De la Rosa et al., 2019): alkyl + *α*-amino C (0–45 ppm); *N*-alkyl + OCH_3 C (45–60 ppm); *O*-alkyl C (60–110 ppm); aromatic C (110–160 ppm); carbonyl C (mainly carboxyl + amide, 160–220 ppm). The ^{13}C intensity distribution was determined by integrating signal intensity over the above-mentioned chemical shift regions. The elemental

analyses of HAs extracted and purified from the soils were conducted on a LECO CHNS-932 instrument.

6.2.3 FTICR-MS analysis

A Bruker 12 Tesla Apex-Qe FTICR-MS instrument equipped with an Apollo II electrospray ionization (ESI) source, operating in negative ionization mode was used. For this analysis, 0.5 mg of purified HA was dissolved in 1% NH_4OH, and a blank sample without HA was prepared. Before injection, the sample was diluted with methanol:water 1:1 v:v to improve the ionization efficiency. The injection was in a flow rate of 120 $\mu L \cdot h^{-1}$ with a nebulizer gas pressure of 20 psi and a drying gas pressure of 15 psi. Peaks identified in the blank were subtracted from the sample peak list prior formula assignment. The empirical molecular formulas were assigned in the range from 200 to 800 m/z, using an in-house Matlab code (The MathWorks, Inc., Natick, MA) according the following criteria: $^{12}C_{2–50}$, $^{1}H_{5–100}$, $^{14}N_{0–6}$, $^{16}O_{1–30}$, $^{32}S_{0–2}$ and $^{32}P_{0–2}$ within an error of 1 ppm, and using the rules outlined by Stubbins et al., 2010. The parameter number of double bond equivalents (DBE), which represent the number of double bonds in a structure and is calculated according:

$$DBE = \frac{1}{2}(2C + N - H + 2)$$

The mass calibration was carried out based on naturally present fatty acids (Sleighter et al., 2008). In order to use appropriate data matrices for chemometric approaches, only molecular formulas identified in all HAs from the different soils were selected. For the selection of the common compounds, after the correct assignment of the formula, the exact mass of the different molecules was used, due to the high resolution of the FTICR-MS, which allows differentiating them. After selecting the common compounds, their intensities were normalized as total abundances i.e., percentages of the total intensity. These compounds were represented in van Krevelen diagrams using the H/C and O/C ratios calculated from each empirical formula. Additional refinement is addressed with surface density plots simultaneously showing compound clusters of points in different regions of the van Krevelen diagram, as well as its total abundances, i.e., percentages of the total intensity. With this purpose, and using authors' own ad hoc computer program, the original $z(x, y)$ data were

transferred into a 50 × 50 matrix (suitable to reallocate the 131 compounds represented in the plane defined by the atomic ratios) by an agglomerative manner. When several compounds coincided in the same H/C and O/C range their intensities values were aggregated. From this matrix, an interpolated surface is obtained by applying the moving average algorithm.

The modified aromaticity index (AI_{mod}), which assumed half of O participating in a double bond, was used to identify the aromatic condensed structures (AI_{mod} > 0.67) (Koch & Dittmar, 2006), this was represented by a diagonal line in Fig. 7.1.

$$AI_{mod} = \frac{1 + C - 0.5O - S - 0.5H}{C - 0.5O - S - N - P}$$

6.2.4 Partial least squares regression

Partial least squares regression (PLS) using the ParLeS program (Viscarra-Rossel, 2008) was applied to simulate models to explain the variance of the SOC content (as dependent variable) in terms of descriptors (independent variables) consisting of the 131 common compounds obtained by FTICR-MS. In particular, PLS was used to explore the utility of the common compounds as predictors of SOC concentration. Prior to statistical analyses, the total abundances of the compounds were managed as compositional data and subjected to the convenient centered log-ratio (CLR) transformation (Aitchison, 1986). To select the minimum number of factors or latent variables (LVs) of each PLS model a series of complementary criteria were considered during the cross-validation with the leave-one-out method i.e.; the root mean squared error (RMSE) and the Akaike's (1974) information criterion (AIC). Finally, an additional more rigorous criterion was used to confirm the possible overfitting of the model, consisting of repeating the PLS study using fully random permutation of the SOC values as dependent variables (the models were discarded as overfitted if some significant (P < 0.05) model is also obtained with any random dependent variable.

6.2.5 Assessment of diagnostic compounds

After confirming the significance of the model explaining the SOC values in terms of the abundances of the common compounds, an evaluation about the individual contribution of these compounds in the prediction model is carried out. With this purpose, additional 2D van Krevelen plot was prepared where the variable importance for projection (VIP) scores calculated during the PLS regression was represented in the *z* axis. This plot is useful to compare the contribution of the independent variables to the prediction model. Nevertheless, by definition VIP values are always positive, not informing on whether the diagnostic compounds are those prevailing in the samples with high or low SOC content.

In order to illustrate compositional differences in terms of SOC levels, new van Krevelen plots were prepared representing the average composition of a number of samples. For this purpose, the samples were ordered according the SOC content and classified by quartiles. The average composition of the 131 common compounds from the HAs of soils in the 1st quartile (Q1, samples with highest SOC content) and the average composition of the HAs from soils with C levels in the 4th quartile (Q4, samples with lowest SOC content) were calculated and normalized as total abundances. Finally, a subtraction of the spectroscopic arrays (with the 131 common compounds) from the HA samples of these quartiles was calculated to represent in *z* axis the differences between the compounds proportions. The resulting three-dimensional van Krevelen plot (with positive and negative values) intuitively illustrate the differences in the molecular composition of the HA in the soils according the SOC levels (Almendros et al., 2018; Jiménez-Morillo et al., 2018).

Finally, the Student's test between concentrations of the compounds released from HAs in soils with different SOC levels was carried out to check for significant differences using the Microsoft Excel (2016) function T.TEST(array1,array2,2,3). In this research, differences ($P < 0.1$) were considered in the discussion of the differential features in the composition of the different groups of HAs with contrasting C levels.

6.3 Results

In Table 4 it is shown the high variability in SOC content (17–157 g C·kg^{-1} soil), and the C/N ratio, which ranged between 8.9 and 31. Table 6 also displays large variability in the elemental composition of HA: C was the most abundant element, ranging from 50.5 to 59.1 g·100 g^{-1}) the next was O with values between 31.2 and 41 g·100 g^{-1}. Finally, the other minor elements were H (3.3 to 5.8 g·100 g^{-1}), N (2.9 to 5.7 g·100 g^{-1}) and S (0.3 to 0.9 g·100 g^{-1}). The solid-state ^{13}C NMR spectra (Fig. 6.1) of the HAs showed the proportions of its C-types, corresponding to different functional groups. The aromatic C content in HA presented values between 14.3 to 35.1%. In the case of *O*-alkyl C, the HA sample with the lowest C content had a value of 14.3% while the highest one reached 27.4%. The content in *N*-alkyl + OCH_3 groups showed a low variation (7.9 to 11.8%). The alkyl C region was the major region of the ^{13}C NMR spectrum, with values between 25.5 and 41.5%. Finally, the carbonyl region presented values ranging from 9.4 to 14.8%.

The FTICR-MS analysis identified between 1000 and 4000 different molecular formulas (> 70% of assigned formulas) for each HA (example showed in Fig. 6.2); most of these compounds were composed by C, H, O and N, some of them including S and P in their structures (Table 11). Although thousands of molecular formulas were identified, only up to 131 compounds, with molecular weight ranging between 309 and 555 Da, were systematically presented in all HAs. These common compounds consisted mainly of C, H and O atoms, only three of these compounds included N. When the common compounds were represented in a van Krevelen diagram (Fig. 6.2) according the H/C and O/C ratios, it was possible to observe that these common compounds can be classified into four groups: two major groups either corresponding to lipid- or lignin-derived compounds, and other two groups with comparatively low number of molecules in regions characteristic for condensed aromatics and protein-derived structures.

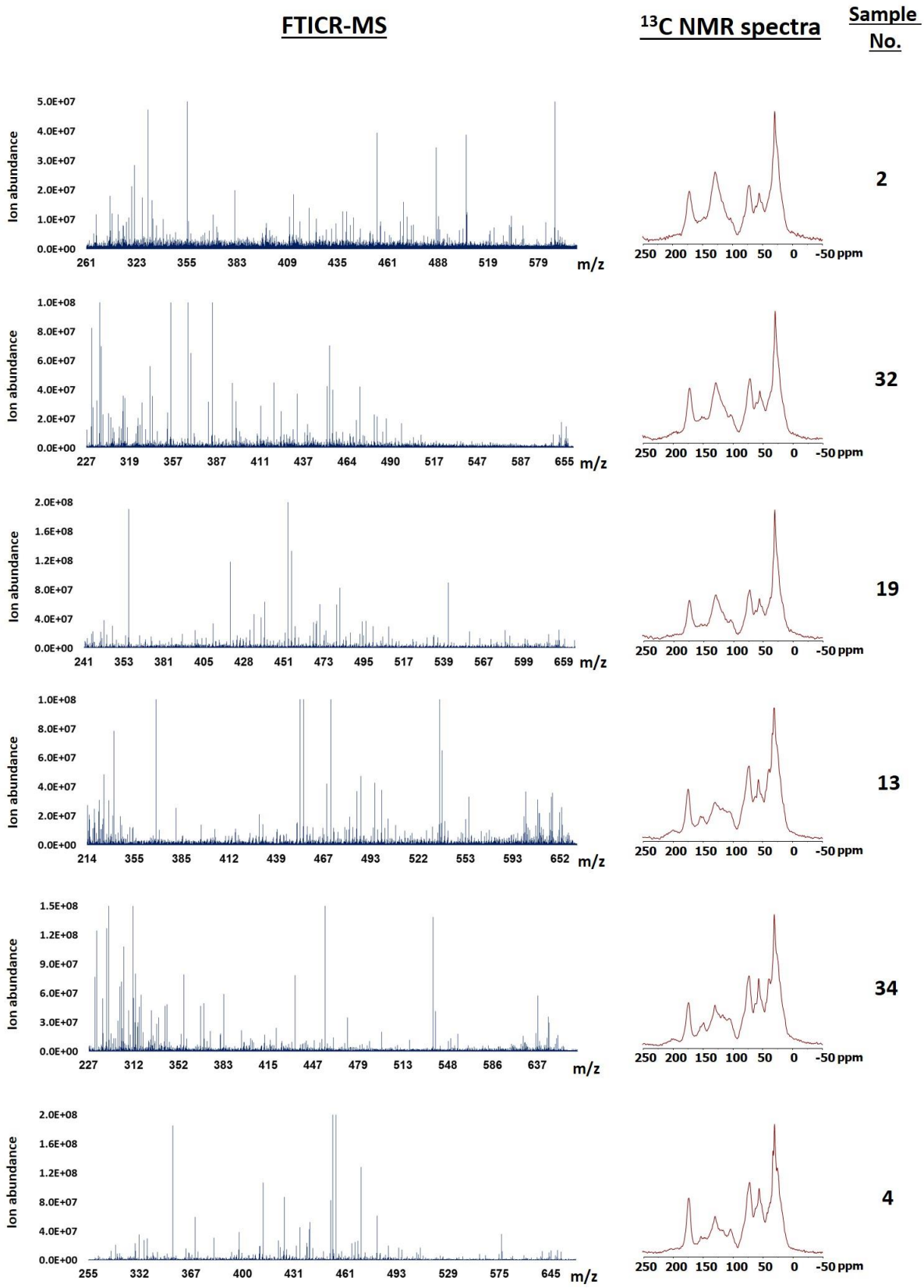


Figure 6.1. Examples of Fourier transform ion cyclotron resonance mass spectra (FTICR-MS) showing ions and nuclear magnetic resonance ^{13}C (NMR) spectra of the corresponding samples of soil humic acids (HAs).

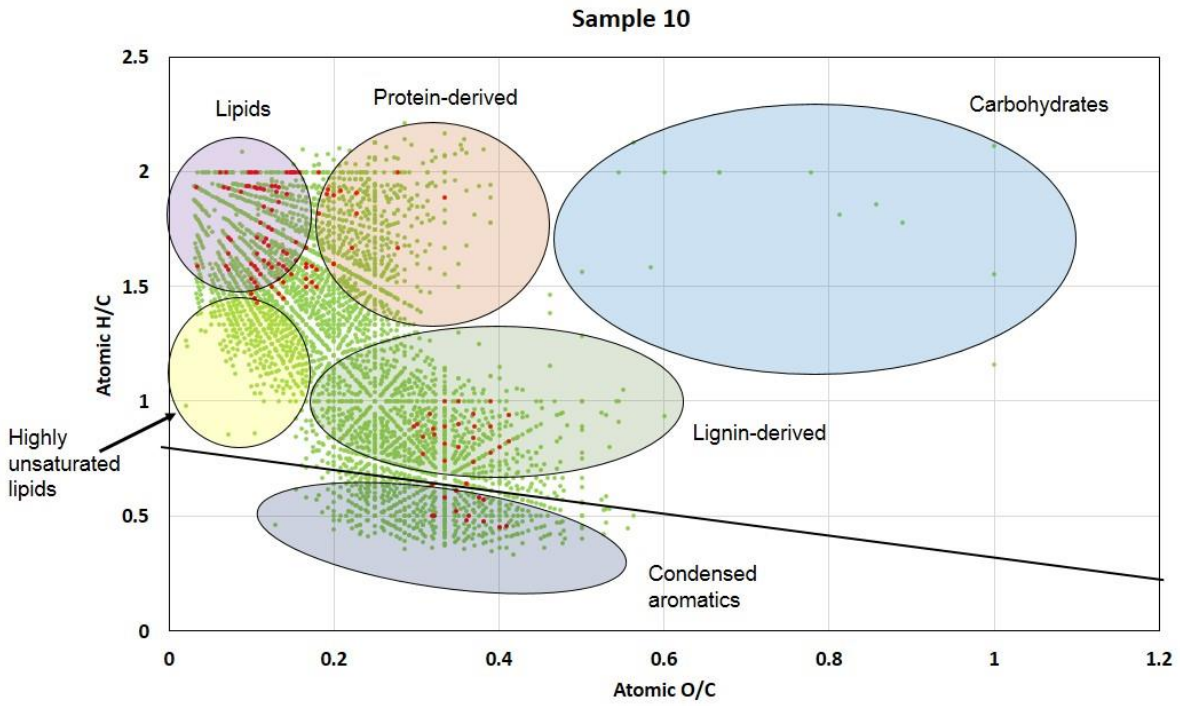


Figure 6.2. Van Krevelen diagram of total compounds detected by Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) in sample 10 (3487 compounds) with the 131 common compounds identified in all the humic acids (HAs) of different soils marked in red. The regions defined by the characteristic H/C and O/C atomic ratios of different chemical structures (lipid, protein- and lignin-derived, carbohydrates and condensed aromatics) are indicated in the diagram. The black solid line denotes boundline of modified aromaticity index (AI_{mod}) = 0.67 (condensed aromatic molecules are below the line).

The comparison of the cross-validation plot obtained in the PLS model from the experimental SOC values with that using fully randomized SOC values were shown in Fig. 6.3. In the best model the values of RMSE and AIC suggested selecting up to 5 LVs to generate the model. In the case of randomized values, the RMSE and AIC calculated with the same number of LVs did not present a progressive trend of the curve (Fig. 6.3 b,d), and the predicted vs observed SOC sets were not correlated. This indicates the reliability and lack of overfitting of the prediction model. Figure 2 shows the observed vs predicted values of SOC obtained using as independent variables the common compounds identified in the HAs. In the cross-validation plots for the PLS model using the experimental values of SOC there was a significant correlation ($P < 0.05$) between predicted

and observed values, with R^2 = 0.6799, whereas in the models using randomized values there was no significant correlation (R^2 = 0.0084).

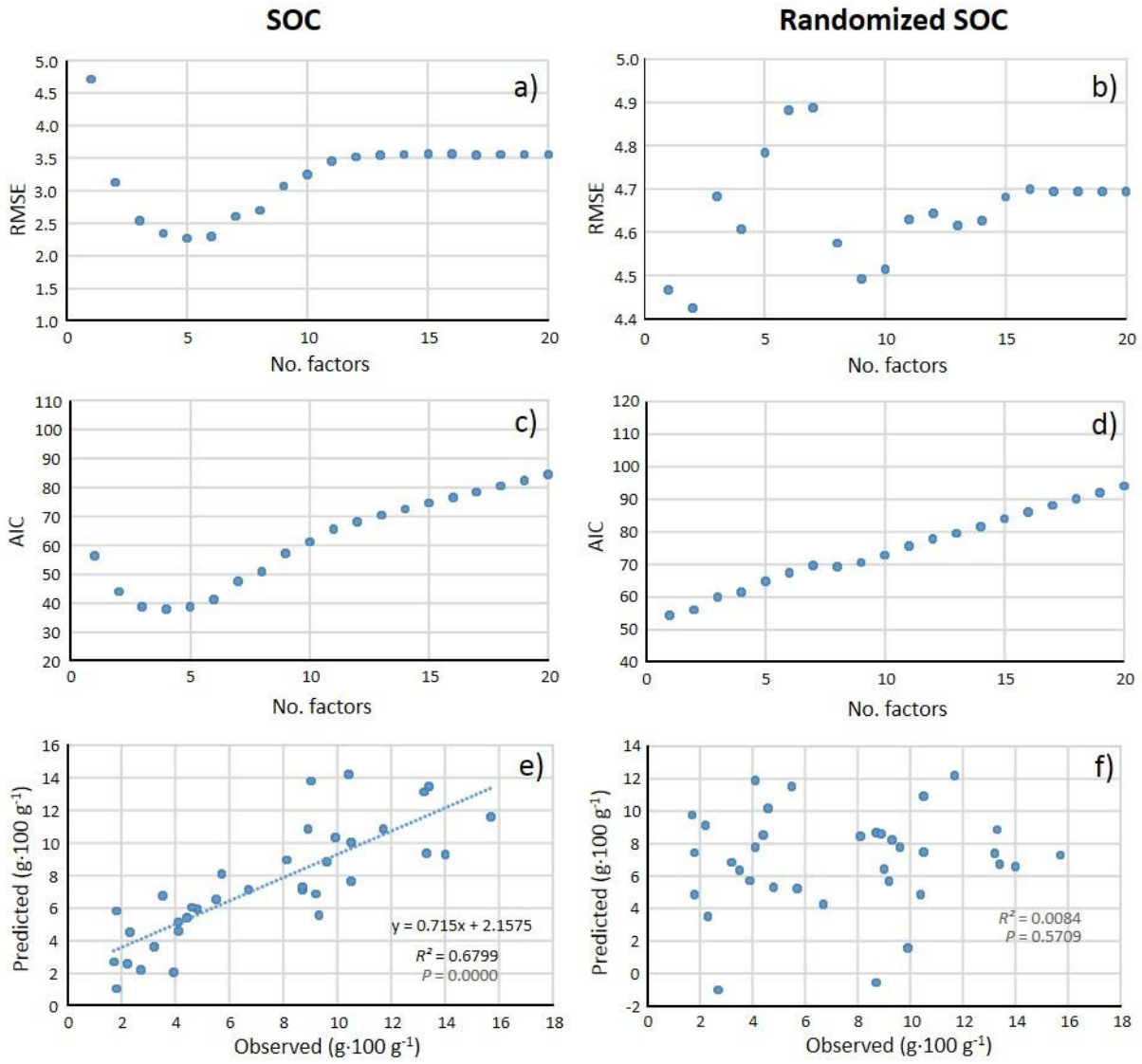


Figure 6.3. Cross-validation plots (experimental vs predicted values) corresponding to partial least squares (PLS) models to predict the concentration of soil organic carbon (SOC). Comparison of results of experimental values of SOC (a, c, e) and randomized SOC values (b, d, f). The root mean squared error (RMSE) of SOC (a) and randomized SOC (b) respectively. The Akaike's information criterion (AIC) (c, d). Observed SOC values vs predicted values obtained by the PLS prediction model for SOC (e) (R^2 = 0.6799) and randomized values of SOC (f) (R^2 = 0.0084) using 5 latent variables or factors suggested by the RMSE and AIC values.

The Fig. 6.4 shows VIP values for each compound in the SOC prediction model. These values inform on the extent to which the different compounds contribute to explain SOC levels. According to the molecular mass (Fig. 6.3a), it was possible to observe how comparatively higher VIP values correspond to molecules with high molecular weight, the highest values corresponding to compounds with molecular weight between 425 and 555 Da. The Fig. 6.4b shows intensity of VIP values for the compounds represented in the van Krevelen diagram. This plot illustrates how the major VIPs correspond to compounds that were located in the lipid region (H/C > 1.4). Two well-differentiated subgroups were observed in the lipid region, these two subgroups can be classified into saturated (H/C > 1.8; DBE < 3) and unsaturated lipids (H/C= 1.4–1.8; DBE= 4–9).

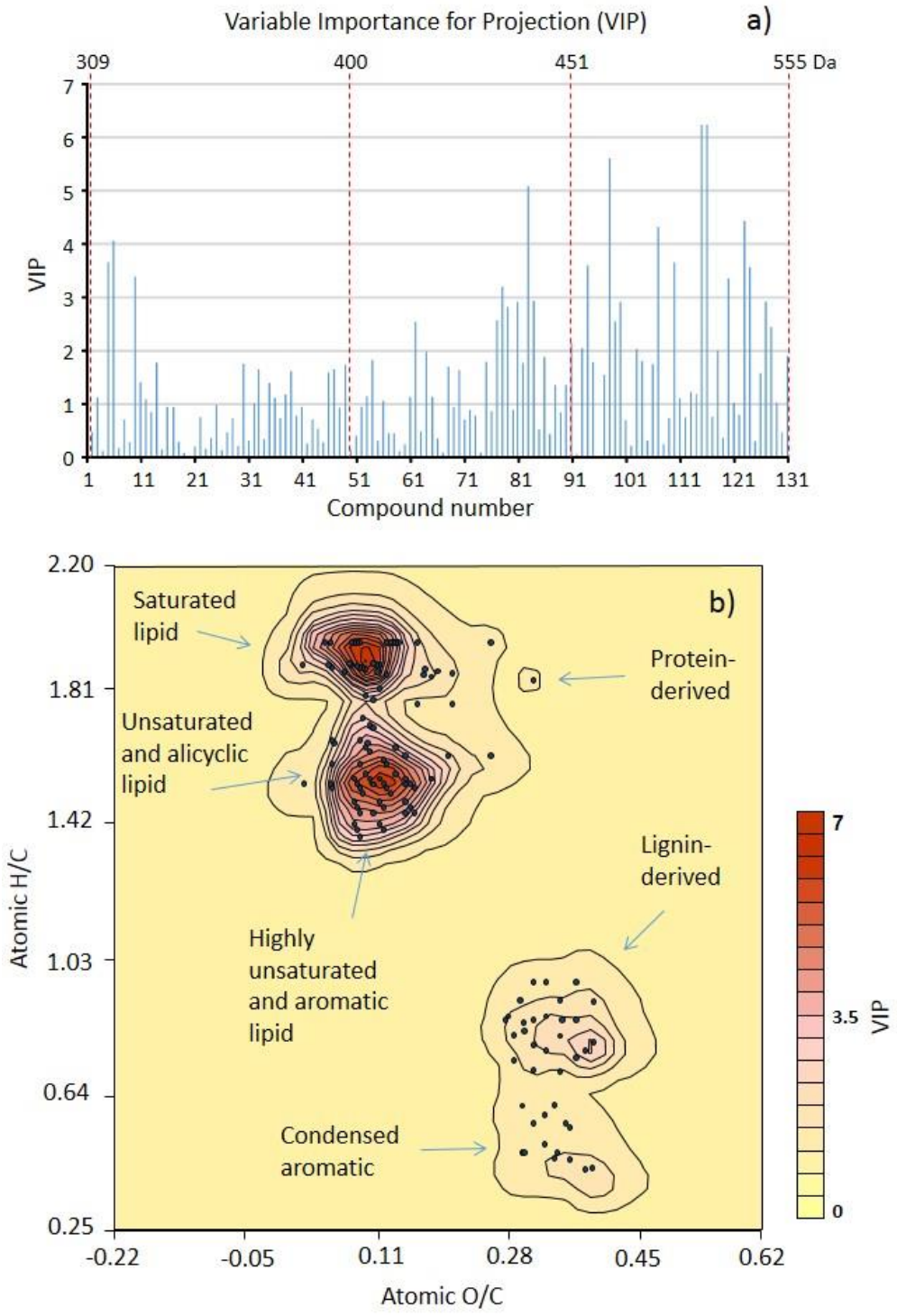


Figure 6.4. Variable importance for projection (VIP) for the 131 common compounds used as descriptors in the partial least squares (PLS) model to predict the soil organic carbon (SOC) content (ordered by molecular weight in *x* axis) in a); van Krevelen diagram simultaneously showing the location of the common compounds and the VIP values shown with colour intensity as a contour diagram in b).

An overall representation of the compositional differences between HAs from the soils representing C-sinks as regards to those with low SOC levels is obtained by subtracting the average values for the total abundances of the individual compounds present in the molecular assemblages of the sample sets with high and low content of SOC i.e., quartiles (Q1-Q4) as shown in Fig. 6.5. This plot shows up to three independent clusters in the lipid region (i.e., saturated (H/C: 1.8–2 and O/C: 0–0.15), unsaturated/alicyclic (H/C: 1.4–1.8 and O/C: 0–0.15) and highly unsaturated/aromatic lipids (H/C: 0.7–1.4 and O/C: 0–0.15). Using the values of the Student's test, the significant ($P < 0.1$) differences between the proportions of the 131 compounds in groups with high and low SOC content were calculated. The resulting values were plotted in the same van Krevelen diagram as a contour diagram superimposed on the values of the corresponding abundances. The Student's *t* test values indicate that the lipid region and the region for lignin-derived compounds are those which showed significant statistical differences between the composition of HAs from soils with C levels in quartiles Q1 and Q4. In particular, the region including compounds with atomic ratios similar to those of lignin structural units showed high values in the HA samples from soils with high SOC content. In the lipid region, it was observed that HAs from soils with low SOC content were comparatively richer in saturated lipid structures. Conversely, unsaturated lipid structures showed the highest values in HAs from soils with high SOC content.

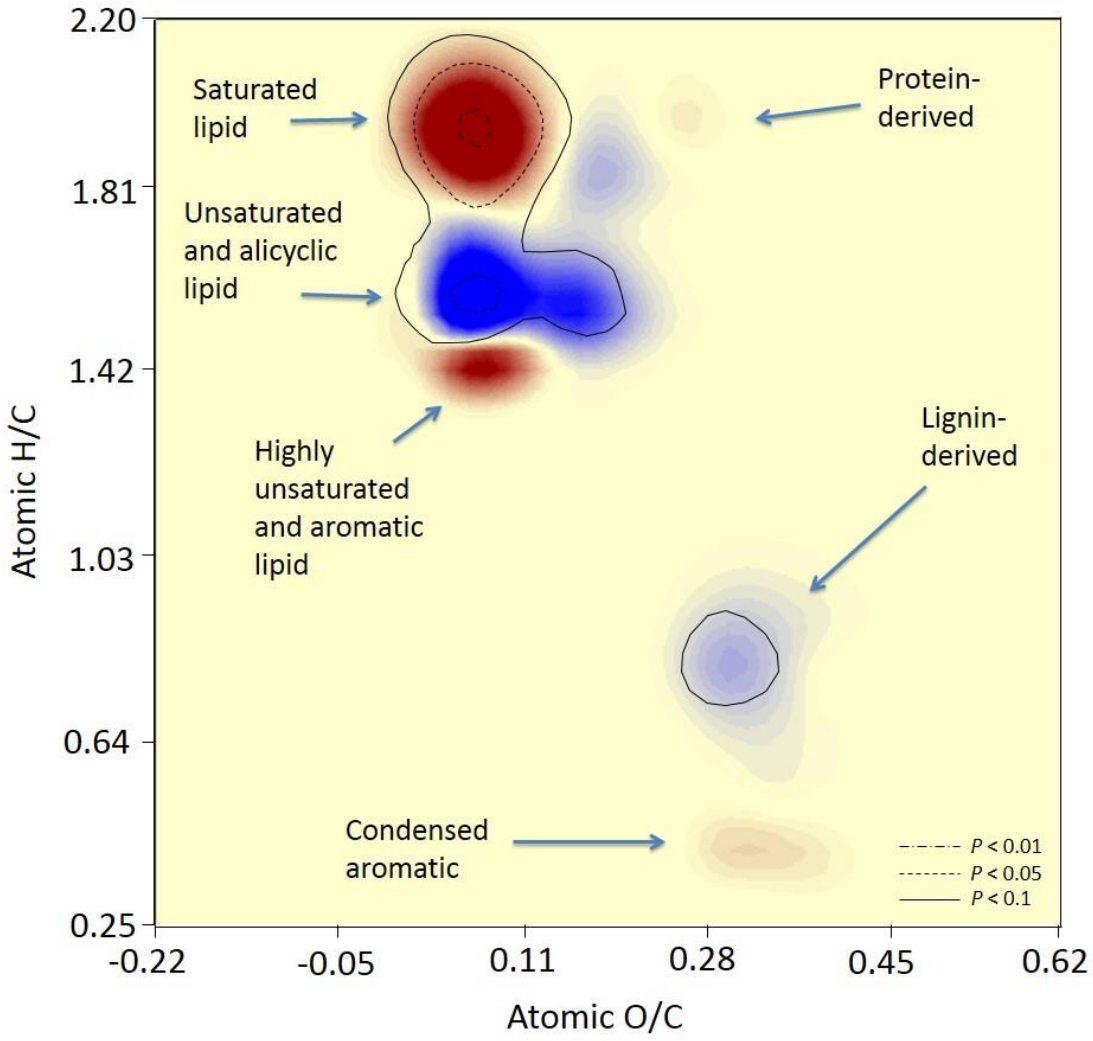


Figure 6.5. Comparison of abundances of common compounds obtained by Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) in terms of the soil organic carbon (SOC) levels of the corresponding soils. The van Krevelen diagram shows the subtraction values between average compounds composition in samples from soils at extreme quartiles as regards the SOC content. Q1–Q4 values with positive values are shown in blue (Q1, high SOC level) and negative values in red (Q4, low SOC content). The superimposed contour map shows the Student's *t* test values showing values statistically different ($P < 0.1$) between the proportions of the molecular fórmulas in the soils with comparatively high C storage potential as regards the comparatively C depleted soils.

6.4 Discussion

Although the HAs showed very similar composition as regards its elemental composition, the analysis by ^{13}C NMR showed differences in the chemical

composition. In particular, ^{13}C NMR regions assigned to aromatic C, which is often considered as an indicator of the progress of the humification process (Tinoco et al., 2014), presented a large variability. This fact indicates variances in the humification process and the resulting HA maturity that are effectively reflected in the quantitative composition of the different C-types in the HAs.

The fact that only 131 compounds were common in all HAs, may be related to the great variability of the soils studied, developed under contrasting formation factors. In any case, the exclusive use of common compounds has the additional advantage of acting as a convenient data reduction method suitable for multivariate data processing, based on a data matrix with an improved normal distribution and no missing values. In the van Krevelen diagram of common compounds (Fig. 6.2), most molecules were located in the regions for lipid and lignin-like compounds. The important extent of the lipid domain illustrates a significant preservation of aliphatic structures, which increase in terms of the humification degree (Stevenson, 1994; Jiménez-González et al., 2017, 2018; Tinoco et al., 2018). On the other hand, the information supplied by lignin-derived compounds suggests a major molecular domain consisting of structural units of biomass from vascular plants, a feature in common with fresh organic matter at early transformation stages and continuously deposited in the soil by vegetation inputs (Fernández-Getino et al., 2013; Miralles et al., 2015; Jiménez-González et al., 2019).

The principal goal of this work is to examine for possible relationships existing between carbon content of the soils and the chemical composition of the corresponding HAs, using the PLS as an exploratory chemometric data treatment. The significant PLS prediction model (Fig. 6.3) obtained for SOC content showed that a relationship exists between the SOC levels and the composition of the HA as reflected by the 131 common compounds analyzed by FTICR-MS. It is probable that this chemical composition and distribution of the abundances of diagnostic compounds as regards soil C storage potential is reflected, on the one hand, by the presence of comparatively biodegradable products of microbial metabolism and, on the other hand, by inputs of fresh organic matter from vascular plants.

It is observed that high VIP values for the 131 individual compounds showed a systematic preference for some type of molecules. In particular, compounds with comparatively higher molecular weight (> 425 Da) displayed higher VIP values

(Fig. 6.4a), suggesting that they play an important role for the prediction model. The fact that the most diagnostic compounds in forecasting the SOC levels selectively correspond to those with high MW could suggest the existence of a threshold of complexity in the FTICR-MS biomarkers to retain sufficient structural information to behave as chemometric proxies of emergent soil properties such is its potential for C sequestration. As a whole, when comparing the chemical structure of the diagnostic molecules (Fig. 6.4b) located in the van Krevelen diagram, high VIP values are located in the lipid region. These compounds could be considered as surrogates for biodegradability and humification stages and may represent a repository of diagnostic molecules informing both on the structural constituents of higher plants (epicuticular waxes, plant tissues…) or microorganisms involved in soil C storage (Nip et al., 1986; Gocke et al., 2013).

Finally, the whole results of the PLS model support the idea that the composition of HAs still retain relevant information of soil properties after alkali extraction, as evidenced by the relationships found with SOM formation processes. This is summarized in Fig. 6.5, showing the pattern resulting from the subtraction of compound abundances of the two average spectra from HAs in extreme quartiles on the SOC distribution. The plot showed the compound groups most characteristic of the structure of HAs from soils behaving as C sinks (blue colour) compared to those that showed greater concentration in the soils comparatively behaving as hot spots, the most significant differences according to Student's test correspond to the abundances of structures derived from lipids and lignin. This corroborates the previous differences in the VIPs calculated by PLS, and suggested that these two regions in the van Krevelen diagram are the most diagnostic source of HA compositional indicators informing on the SOC potential storage. As a whole, the predominance of unsaturated and alicyclic lipid was observed in HAs from soils with high SOC content, whereas mainly saturated alkyl molecules were in characteristically high proportions in the HAs from soils of low C content. In the case of lignin-derived compounds, its predominance is indicative of HAs from soils with high content of SOC; similar results were found by Jiménez-González et al. (2019). This suggests that these compounds, indicating recent input of plant material, are useful to explain to large extent the variability of the SOC levels in our soils (Jiménez-González et al., 2017).

6.5 Conclusions

Chemometric assessment based on compounds detected by FTICR-MS depicts the HA structure as a fingerprint for SOC storage processes. A significant relationship exists between the SOC levels and the molecular composition of HAs studied by FTICR-MS, to the extent that it is possible to obtain a very significant prediction model using only the common compounds released by FTICR-MS from the HAs. The fact that the structural features of the HAs are linked to the SOC content of the corresponding soils indicated that HAs ought not to be considered artefacts generated in the alkali extraction, because its composition still reflects biogeochemical information relevant about soil emergent properties as the SOC storage even after structural changes inherent to its isolation by alkali extraction.

The information window provided by FTICR-MS could be considered a valuable source of surrogates in the form of compounds reflecting the effects of the environmental factors controlling the different SOC levels. Soils with high levels of SOC content accumulate HAs rich in unsaturated lipid and lignin-derived compounds, whereas HAs in soils with low SOC content present a composition in which saturated lipid compounds predominate.

Capítulo 7:

Factors of soil formation, effects on soil organic matter structure and soil carbon storage

Este capítulo ha dado lugar a un artículo que se encuentra en proceso de revisión: Jiménez-González M.A., Álvarez A.M., Carral P., Almendros G. Factors of soil formation, effects on soil organic matter structure and soil carbon storage.

Abstract

There is currently an active controversy about the variable influence of the factors involved in the total content and quality of the soil organic matter (SOM), which translates into its resilience and stability against biodegradation, and importantly on the rates of release of CO_2 into the atmosphere. The aim of this work is to study the molecular composition of SOM in contrasting environments in order to evaluate the extent to which such conditions may affect SOM characteristics in addition to the levels of soil organic C (SOC). Up to 33 soils from different environmental scenarios of Spain were analysed by pyrolysis combined with gas chromatography mass spectrometry (Py-GC/MS). The 193 major pyrolysis compounds released from the soils were included in a chemometric study based in discriminant analysis to assess the impact of classical soil forming factors (i.e., climate, vegetation and geological substrate) in SOM content and composition. Improved van Krevelen diagrams are used to facilitate the recognition of different patterns in SOM composition dependent on soil forming factors. The results show different patterns in the SOM composition depending on its soil forming factors, with a decreasing influence in the order: climate > vegetation > geological substrate. In addition, the total levels of SOM were also different depending on the environmental scenarios on these soils, suggesting both qualitative and quantitative control of soil carbon sequestration.

7.1 Introduction

The factors involved in the transformation of soil organic matter (SOM) are extremely complex and responsible of its total levels and quality. While some authors highlight the importance of climatic factors in controlling the soil formation processes, others consider the effect of vegetation or geological substrate more important (Duchaufour & Jacquin, 1975; Ganuza & Almendros, 2003; Jenny, 1994; Johnson et al., 2011; Towett et al., 2015). In addition, there is still no consensus regarding the quantitative influence of the factors that determine the fact that some soils store more organic carbon than others, which results in important differences in their potential for sequestration of atmospheric carbon that is currently a subject of great interest (Lal, 2004). At this respect, it is also well known that organo-mineral interactions and microencapsulation of the SOM may play an important role on its physico-chemical protection (Simonetti et al., 2017; Song et al., 2014; Spaccini et al., 2002). On the other hand, land use may also affect soil organic carbon (SOC) sequestration processes (Hernández et al., 2019; Pizzeghello et al., 2017; Yazdanshenas et al., 2018). Apart of this and regarding soil carbon storage, it is also relevant to point out that the chemical composition of the SOM may also be an important constraint in its recalcitrancy (Jiménez-González et al., 2017, 2018, 2019). This is traditionally taken into account when reffering the melanization of soil horizons, which reflects the fact that, depending on the environmental conditions, the accumulation of humus is variable in quantity and quality (Bockheim & Hartemink, 2017; Di et al., 2019).

In this context, the use of analytical degradation techniques for the molecular characterization of the SOM, such as Py-GC/MS, may represent a source of semiquantitative data suitable to explore the extent to which the composition of the SOM may represent a reliable repository of environmental information about the influence of soil forming factors. This technique has a series of advantages such as not requiring chemical pretreatments, or previous SOM isolation with alkaline reagents. Consequently, a major goal of this research is to assess the influence of the above factors on the humification processes but also to assess the biogeochemical meaning of the variability in the molecular composition of the SOM in terms of the different SOC levels of the soils.

7.2 Materials and Methods

7.2.1 Study area

Topsoil samples (0–10 cm) were collected from thirty-three ecosystems from Spain with a large variability in local climatic conditions, vegetation and geological substrate. The soils were previously classified according to the IUSS Working Group WRB (2014) system (Table 1). The sampling sites were chosen to cover a large range in SOC content. Three sampling points were selected for each sampling area, the soil samples were collected from the A horizon after removing the litter layer (1–5 cm, depending on the morphology of the soil profile). In a second stage, a series of composite soil samples were prepared by mixing samples collected from three different points, and then the samples were air-dried and sieved (< 2 mm).

The soil textural analysis was carried out by using the densimeter method (Bouyoucos, 1927). The pH was determined in soil-water suspension (1:2.5, w:w) with an XS pH meter model pH 7 (Carpi, MO, Italy). A 1*M* ammonium acetate solution at pH = 7 was used to determine the cation exchange capacity (CEC) according to Juo et al. (1976). The SOC concentration was determined by wet chemical oxidation with 1*N* potassium dichromate (Nelson & Sommers, 1982; Walkley & Black, 1934) and the N by micro-Kjeldahl digestion (Prince, 1945).

7.2.2 Analytical pyrolysis

The molecular composition of the SOM was analyzed by Py-GC/MS using whole soil samples. Homogenized soil samples were prepared from 5 g samples grounded to fine powder (< 0.01 mm) with a planetary agate ball mill.

Pyrolysis was carried out at 500 °C with a PY-2020iD pyrolyser (Frontier Lab Ltd., Fukushima, Japan) coupled to an Agilent 6890 GC/MS system with a phenylmethylsiloxane column (Agilent HP-5MS 5%). Helium was used at a flow rate of 1 cm^3 min^{-1} as carrier gas. The GC oven temperature was set to 50 °C for 1 min, then increased to 100 °C at 30 °C min^{-1}; from 100 to 300 °C to a rate of 10 °C min^{-1}, and finally isothermal at 300 °C for 10 min. An Agilent 5973 quadrupole mass spectrometer detector was used, and mass spectra were

acquired using 70 eV ionizing energy. The peak areas (total area counts) in the chromatograms were integrated for the different compounds and expressed as total abundances. Finally, cumulative values of the main families of compounds (viz. alkanes (A), olefins (O), fatty acids (F), phenols (P), methoxyphenols (M), N-compounds (N), alkylbenzenes (B), polycyclic aromatic hydrocarbons (H), carbohydrate-derived compounds (C) were calculated to be used as independent variables in the discriminant analysis.

7.2.3 Statistical analysis

The influence of the different factors on the SOM composition was explored using discriminant analysis (Fig. 7.1). In this treatment, the relative abundances of the different families of pyrolytic compounds were used as descriptors (independent variables). On the other hand, the dependent variables (classification pedogenesis factors) consisted of the different states that define the main soil forming factors (vegetation type, geological substrate and climatic conditions). These different qualitative sample descriptors for each of the soil forming factors were, for the geological substrate: igneous, metamorphic and sedimentary rocks. For the vegetation: Cupressaceae, Fagaceae and Pinaceae species and for the climate the Köppen classification (Kottek et al., 2006) was followed: Cfa, Cfb and Csa types. In a second stage, after performing the discriminant analysis and considering the results in the classification table, the soil samples incorrectly classified were discarded, then the average composition of the different 193 pyrolytic compounds was calculated for each soil group using only the samples correctly classified in terms of the results of the discriminant analysis. This approach led us to obtain a series of numerical arrays corresponding to average pyrograms of SOM samples formed under each factor, in order to subsequently prepare plots based on the classical van Krevelen diagram where these compositions of SOM are represented. Finally, the average SOC content was also calculated for the same groups of samples and the statistical significance of the differences between the average SOC values of the groups was checked at 90% ($P < 0.1$) using the Student's *t* test.

The compounds identified by analytical pyrolysis were essentially the same for all samples but with quantitative differences in their abundances between soils. For a simplified perceptual interpretation of the results we used plots based in the classical van Krevelen (1950) graphical-statistical method. This procedure

has been extensively used to represent the elemental composition and transformation reactions of complex macromolecular materials such as coals or humic substances (Ikeya et al., 2015), because of its potential of intuitively displaying the molecular composition of samples under study. This plot has proven to be useful to facilitate the interpretation of the complex assemblages of individual compounds released by different analytical methods (Almendros et al., 2016; Ikeya et al., 2015; Kramer et al., 2004). Essentially, the approach used for this study consists of plotting "surface density plots" built from the abundances of the individual pyrolysis compounds represented in the space defined by their H/C and O/C ratios (Almendros et al., 2018). For this purpose, the scores for atomic O/C and H/C ratios of the individual pyrolysis compounds are represented in the basal plane (*x,y* axes, i.e., the classical van Krevelen diagram). The vertical dimension (*z* axis) corresponded to the normalized abundances of the individual compounds (sum= 100). Using authors' own ad hoc computer program, the original *z* (*x, y*) data were transferred into a 50 × 50 matrix (suitable to reallocate the 193 individual compounds represented in the plane defined by the atomic ratios) by an agglomerative manner. When several compounds coincided in the same H/C and O/C range their abundances were aggregated, i.e. the case of olefins. From this matrix, an interpolated surface is obtained by applying the moving average algorithm (i.e., averaging each cell value with those of its 4 orthogonal neighbour cells). The resulting plot shows a series of broad 3D peaks or compound clusters, whose size is proportional to the collective value of the abundances of the compounds with similar elemental composition and chemical nature (alkanes, olefins, phenols, etc). This method allows us, in a very perceptual way, to observe the different amounts of compounds which tend to accumulate in the SOM formed under the influence of the different factors (Almendros et al., 2018). In principle, the density surfaces, as the original van Krevelen plots (Fig. 7.2), are valid only to display the molecular composition of individual soil samples, which limits its usefulness to discuss differences among the 33 different van Krevelen graphs. For this reason, here we build up van Krevelen surfaces representing the average compound composition of a set of pyrograms from samples sharing a defined common characteristic i.e., the same environmental characteristic. Further refining is based on comparing additional plots representing subtraction values between the average pyrograms from soils developed under each environmental factor and the total average of the 33 pyrograms, which are

shown as van Krevelen surfaces. These subtraction plots are useful to display the differential characteristics between the pedogenesis factors, showing positive or negative peaks depending on whether the corresponding compounds predominate or are in comparatively low proportions in the defined soil groups.

Finally, using the Student's *t* test for each compound, it is observed whether the above differences between compounds proportions in the soil groups developed under different formation factors are considered significant or not. In this work, differences with $P < 0.1$ have been represented in the plots.

7.3 Results

7.3.1 Chemical analysis

The general analytical characteristics of the soils (pH, CEC, SOC) showed large variability between the samples studied (Table 4). Soil samples showed pH values ranging between 3.9 and 7.7 and a CEC values between 4.5 and 41.9 $cmol_c \cdot kg^{-1}$. The SOC content presented a wide range (18–157 $g \cdot kg^{-1}$). With the Py-GC/MS analysis it was possible to identify up to 193 different compounds in the whole soil samples (Table 10). The semiquantitative results for the pyrolytic compounds normalized as percentages (considering 100% the sum of all peaks identified) were grouped into different families and used as descriptors to perform the discriminant analysis from these reduced variables set. Table 12 shows that the main families of pyrolytic compounds showed a large variability in their total abundance between the soils. The proportions of methoxyphenols, which are typical pyrolysis products from lignin, ranged between 0.1 and 27.8 %. Other aromatic compounds as phenols and alkylbenzenes were also abundant with percentages of 5.1–17.7% and 11.1–46.4%, respectively. Aliphatic compounds included alkanes (0.1–17.0 %), olefins (3.3–20.8%) and fatty acids (0.0–4.7%). N-compounds that are related with protein ranged between 6.9 and 21.1%. Polycyclic aromatic hydrocarbons (5.2–18.5%) and carbohydrates (5.1–17.9%) were present in similar levels. Finally, some steroids were identified (0.1–5.8%).

7.3.2 Discriminant analysis

The results of the discriminant analysis showed that it is possible to identify the soil formation factors exclusively using the chemical information provided by analytical pyrolysis. At first sight (Fig. 7.1) it was observed that the different soil groups selected in each forming factor result in disjoint clusters in the plot; only in the case of geological substrate a somewhat higher overlap between the three groups was observed. Based on the number of samples correctly classified (also reflected in the plot as the sharpness of the clusters indicated with the size of the whiskers), it may be considered that most of the classification factors studied (climate and vegetation) allowed to obtain an excellent value of samples correctly classified (93.9% for climate, 91.7% for vegetation) whereas in the case of geological substrate the number of samples correctly classified was comparatively lower (57.6%). When the above results are compared with those of the average SOC content of the groups, it was observed that the differences were significant between soil groups under different environmental conditions for each soil forming factor.

When using climatic types as classification factor, the three groups were found to be significantly different in terms of its SOC content. Soils under Cfa climate had the highest SOC content, followed for Csa and finally by Cfb, with the lowest content. As regards the vegetation type, the differences were also significant ($P < 0.05$) in two of the groups (Cupressaceae and Fagaceae families), which showed significant differences in the SOC: the soils developed under the former vegetation showed higher SOC content that soils under the latter. Concerning the geological substrate, soils on metamorphic and sedimentary substrates also had comparatively high SOC content than soils on igneous rocks.

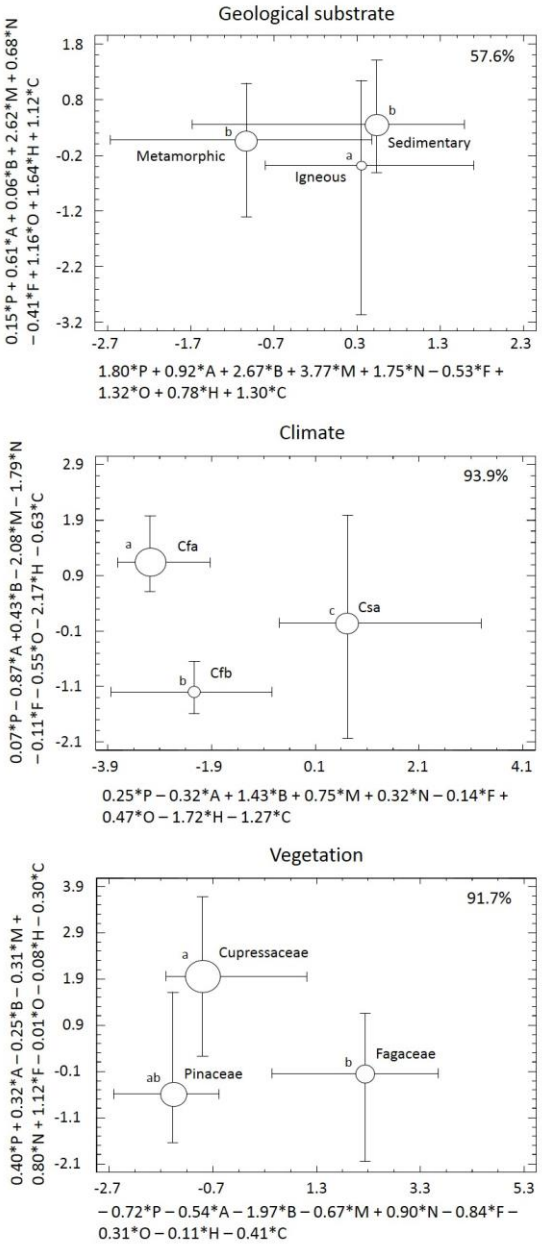


Figure 7.1. Results of discriminant analysis applied to the assessment of the relative importance of different soil forming factors in the molecular composition of soil organic matter and the total soil organic carbon (SOC) content. For each soil group, the position of the centroid (circle) and the variability between the scores of the different soils (whiskers) are shown. Circles area is proportional to the average SOC for each group. The lower case letters close to the centroids indicate soil groups which are statistically different in SOC content according to Student's *t* test at 90% ($P < 0.1$). The descriptors used for this analysis were the proportions (total abundances) of pyrolytic compounds: alkanes (A), olefins (O), fatty acids (F), phenols (P), methoxyphenols (M), N-compounds (N), alkylbenzenes (B), polycyclic aromatic hydrocarbons (H) and carbohydrate derivatives (C). The coefficients of the discriminant functions are shown on the axes. The percentage of soils correctly classified in the different group is shown at the right upper corner (%).

7.3.3 Van Krevelen plots

In Fig. 7.3, the van Krevelen diagrams corresponding to subtraction values between the average abundances of the pyrolytic compounds from soils developed under each factor and the general average from the whole samples studied are shown. The first factor studied was climate: soils under Cfa climate comparatively showed low proportions of alkanes, aromatic constituents (mainly short-chain alkylbenzenes, aromatic N-compounds, polycyclic aromatic hydrocarbons) and methoxyphenols, but comparatively increased proportions of phenols and fatty acids. The following climatic type in terms of SOC content was Csa, which correspond to scores showing similar relative depletion in aromatic compounds including polycyclic aromatic hydrocarbons. Finally, the SOM of soils under Cfb climate, with the lowest SOC content, characteristically had higher proportions of olefins, alkylbenzenes, aromatic N-compounds and polycyclic aromatic hydrocarbons, and lower proportions of methoxyphenols and fatty acids than the average.

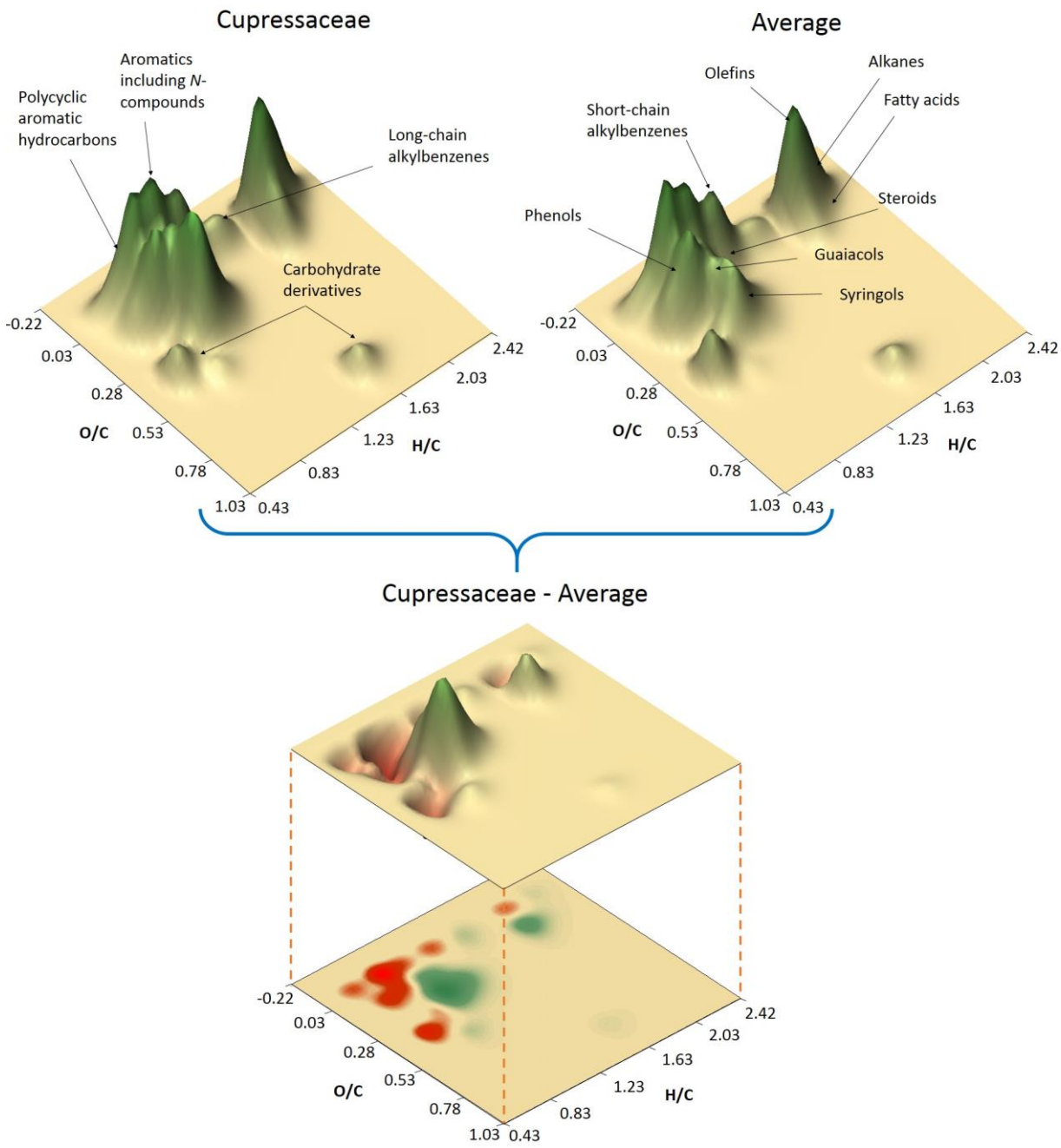


Figure 7.2. Example of subtraction of van Krevelen diagrams in which average values of the abundances of the pyrolytic compounds from the whole soil sample set (33 soils) is subtracted from the average values of the abundances of the same compounds in soils developed under vegetation of Cupressaceae. The subtraction plot, either represented as surface density plot or density map, allows to easily detect the groups of compounds that differ in abundance between the two groups of samples. The compounds that are more abundant in the soils developed under Cupressaceae than in the general average of the 33 soils are shown in green colour, while those present in comparatively lower proportions are shown in red.

Concerning vegetation types, the SOM under Cupressaceae showed comparatively low values of olefins, phenols, long-chain alkylbenzenes, aromatic N-compounds and polycyclic aromatic hydrocarbons. This group showed high proportions of fatty acids, short-chain alkylbenzenes and methoxyphenols, in particular guaiacyl-type. The same occurs in the Pinaceae group: in the subtraction plots, alkanes, olefins, alkylbenzenes, aromatic N-compounds and polycyclic aromatic hydrocarbons showed negative values. Phenols and guaiacyl-type methoxyphenols had positive values, whereas syringyl-type methoxyphenols presented lower values than the average. Finally, the SOM in soils under vegetation of Fagaceae showed slightly lower proportions of aromatic compounds, phenols, fatty acids and carbohydrate-derived compounds. Olefins, polycyclic aromatic hydrocarbons and syringols showed greater values than the average of the whole soil set.

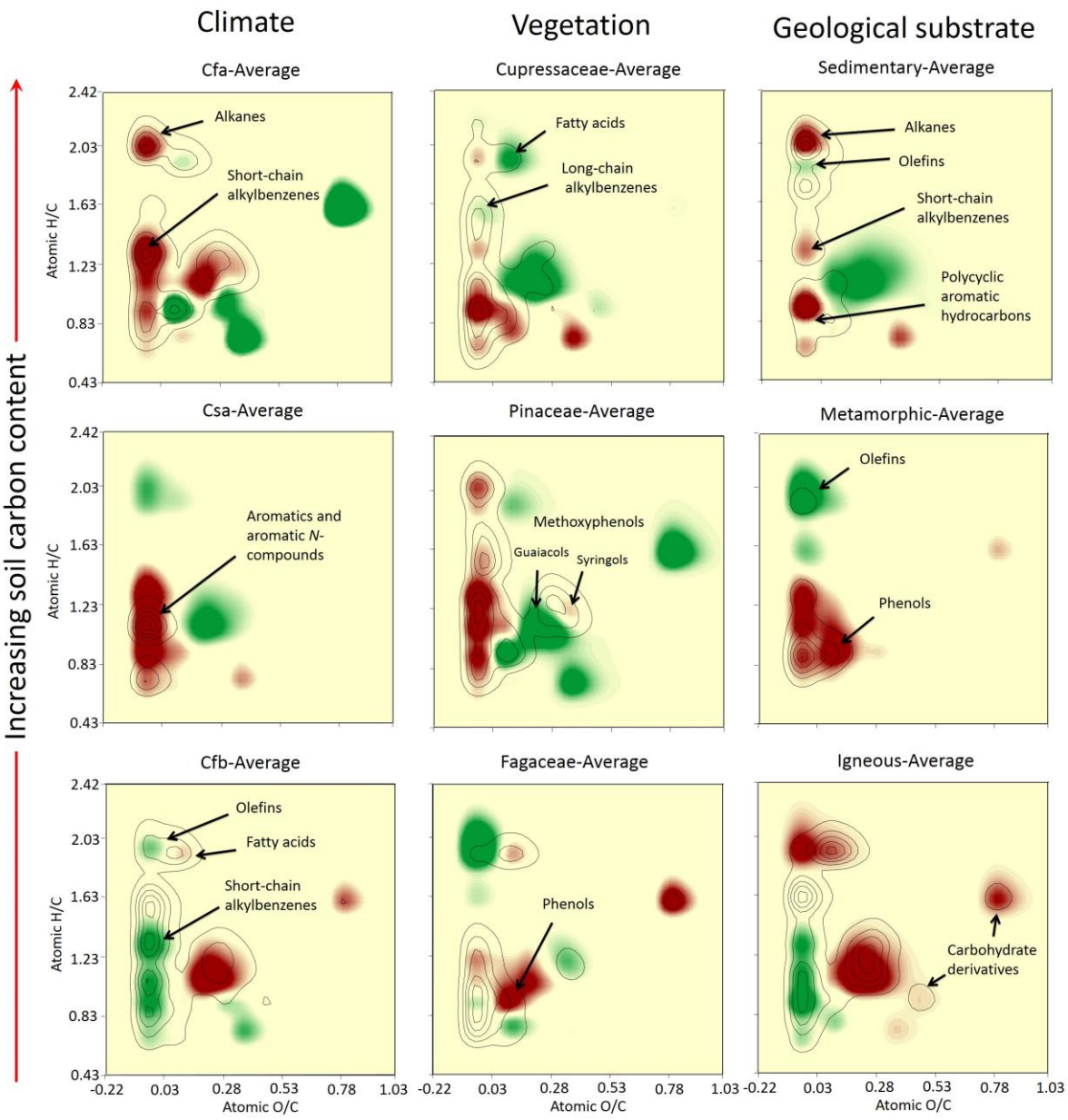


Figure 7.3. Van Krevelen diagrams obtained by subtracting values of the abundances of the pyrolytic compounds from the average of the whole soil sample set (33 soils) minus the average values of the abundances of the same compounds in soil subsets (consisting of the soils developed under different soil forming factors, and correctly classified according to the previous discriminant analysis) and shown as density maps in the space defined by their H/C and O/C atomic ratios: molecular constituents of soil organic matter that prevail in soils developed under the corresponding formation factor (positive values after subtraction) are shown in green, whereas those present in lower proportions than in the general average are shown in red. The superimposed contour diagram shows the significance level of the difference between compounds abundances in the general average and averages from the different subsets corresponding to individual forming factors (i.e., significance levels of the Student's *t* test between each compound in the two soil sets, $P < 0.1$ in the external contour).

Finally, the geological substrate was the classification criterion leading to the lowest number of samples correctly classified by discriminant analysis. The SOM composition in the soils developed under sedimentary rocks released a comparatively low proportion of alkanes, short-chain alkylbenzenes and polycyclic aromatic hydrocarbons, conversely showing positive values for olefins and phenols. The metamorphic substrate is associated with SOM with high proportions of olefins but low of phenols, alkylbenzenes, aromatic N-compounds and polycyclic aromatic hydrocarbons than the average. The igneous rocks were associated to soils with low SOC content; the subtraction values with respect to the whole set of samples displayed a compositional pattern with negative values for olefins, fatty acids, methoxyphenols and carbohydrate derivatives, whereas the proportions of aromatic compounds (alkylbenzenes, aromatic N-compounds) and polycyclic aromatic hydrocarbons were higher than the average.

7.4 Discussion

The application of discriminant analysis demonstrated that the soil-forming factors considered in this research (vegetation, climate and geological substrate) exert a significant influence not only in the total SOC content but mainly in the SOM composition as reflected by analytical pyrolysis of whole soil samples. Climate and vegetation showed the most outstanding effects, whereas the type of geological substrate is reflected to a lesser extent in the molecular composition of the SOM.

Other important variable under examination is the SOC content, which represents an indirect measure of the potential for SOC storage, and was found largely influenced by the soil-forming factors studied. It is possible that the above qualitative and quantitative features of the SOM varying in terms of the forming factors could be causally related, an aspect that would deserve further detailed investigation. Nevertheless, it is clear that the results point to the fact that soils with significantly different C contents have also SOM with significantly different molecular composition, irrespective to the fact that these differences in molecular composition are the cause or the effect or its stability or recalcitrance. This latter aspect is relevant with respect to the fact that soils with high SOC content are not necessarily soils where SOC is more stable or recalcitrant. This

is clearly observed in the composition displayed in the van Krevelen diagrams (Fig. 7.3).

The van Krevelen diagrams illustrated how the SOM composition differs significantly in soil groups classified according to its common soil-forming factors. Considering the differences among the groups in terms of SOC content, the most important difference was the tendency to accumulation of aromatic structures in soils with comparatively low SOC content. This fact has been frequently correlated with SOM recalcitrance, considering that SOM aromaticity often behaves as a surrogate of its chemical stability and resistance to degradation (Miralles et al., 2015; Tinoco et al., 2015). Despite its comparatively low SOC content, this may be more recalcitrant than in other soils with higher SOC values. In fact, C storage is usually carried out at expenses of labile fresh organic matter, that is reflected by the dominance of lignocellulose (polysaccharides and methoxyphenols). A similar situation occurs with the alkanes; the proportions of alkanes showed a clear variation with the SOC content, in agreement with its chemometric potential in forecasting SOC levels described by Jiménez-González et al., (2018).

The differences depending on the geological substrate could be interpreted as the effect of a different nature of the organo-mineral interactions controlling the humification processes (Nezhad, 2019). This fact would also depend on the greater or lesser capacity of weathering of the different types of rocks and the mineralogy of the resulting clay fraction. The most characteristic pattern was observed under metamorphic rocks, associated to the accumulation of SOM types where, as regards the average of SOM composition, there were no significant differences in the proportion of methoxyphenols or carbohydrate-derived compounds, but preferential accumulation of condensed alkyl constituents mainly olefins was observed, whereas aromatic products including phenols do not tend to accumulation. In the case of the sedimentary geological substratum—which includes soils on limestone—no large differences are observed in the methoxyphenol or carbohydrate region of the van Krevelen diagram: although the values of the former compounds may differ from those in the general average, the differences are not highly significant. These soils on sedimentary substrate show low proportions of aromatic compounds, especially of high condensation, which could correspond to a possible encapsulation effect by carbonates or active colloidal minerals which are traditionally considered to

delay the transformation of the SOM and its evolution towards more advanced forms (Duchaufour, 1970). Finally, SOM of soils developed on igneous rocks apparently show a high transformation reflected by the significant depletion of oxygen-containing constituents compatible with an origin from both vegetation and microorganisms (i.e., yielding all types of phenols in addition to lipid- and carbohydrate derivatives). The most characteristic pyrolysis products consisted of alkylbenzenes, aromatic N-compounds and polycyclic aromatic hydrocarbons.

When comparing the SOM composition in soils under different vegetation types, it can be seen that in soils under Fagaceae the composition is compatible with high microbial activity, with accumulation of alkyl compounds. In addition, carbohydrate derivatives and phenolic compounds are degraded comparatively, while the accumulation of low proportions of condensed aromatic compounds is significant. However, under Gymnosperm vegetation, and in particular under pine trees, the SOM is comparatively less transformed, as constituents of lignocellulosic biomass such as methoxyphenols and carbohydrate derivatives predominate. Conversely, pyrolytic products characteristic for higher degree of transformation of plant biomass, e.g. alkylbenzenes, aromatic N-compounds and polycyclic aromatic hydrocarbons do not tend to accumulate in SOM. In general, the patterns in the van Krevelen diagram are quite similar in the two groups of Gymnosperms, Pinaceae and Cupresaceae, but SOM quality would be lower in the case of the Pinaceae attending to the significantly higher proportions of syringyl type methoxyphenols (indicating preservation of the less-condensed, easily biodegradable lignin) and carbohydrate derivatives (Jiménez-González et al., 2017, 2019; Miralles et al., 2015).

When comparing the effect of climate simultaneously considering differences in moisture and temperature, it would be possible to describe the different SOM evolution depending on whether, the climate tends to greater tropicality or to desertification. For instance, the coldest climatic type Cfb favour the accumulation of organic matter composed mainly of non-methoxylated aromatic compounds, in addition to alkyl and fatty acid type constituents, the proportions of which being significantly higher as regards warmer climate types, Cfa and Csa (Fig. 7.3). In general, under warmer climatic conditions (both Cfa and Csa) the accumulation of aromatic, polycyclic aromatic hydrocarbons and alkyl constituents in the SOM is to lesser extent favoured, probably due to the

historical influence of wildfires (Jiménez-González et al., 2016) which have a high frequency in this climatic conditions. Concerning the simultaneous increase in moisture and temperature, i.e., a tendency to tropicality that would correspond to an evolution from Cfa climate to Csa, Fig. 7.3 suggests changes in SOM consisting of a comparatively lower accumulation of alkyl compounds than in warmer and dry conditions (Csa). In the case of differences more typical for a trend towards desertification (from Cfb to Csa), the soil showed accumulation of raw humus, where lignin-derived methoxyphenols prevail on condensed aromatic compounds. Considering the narrow climatic range amongst all soils studied under Mediterranean ecosystems, it seems clear that a progressive warming would lead to lower quality of the SOM at least defined to the extent that the constituents of the biomass are preserved and are not replaced by more or less condensed aromatic products and fatty acids compatible with microbial reworking processes.

7.5 Conclusions

There is a clear distinction in the molecular composition of SOM as a function of the impact of different soil-forming factors. It was found that the transformation processes of SOM in the topsoil differ significantly in terms of vegetation, climate and, to a lesser extent, of geological substrate. In addition, significant differences in SOC storage depending on the environmental characteristics studied were also observed. For instance, soils under Gymnosperm vegetation or developed on sedimentary geological substrate, which have high levels of SOC, display soil C storage processes mainly at expenses of preservation of slightly transformed lignin. Concerning SOM quality, soils developed under Cfb climatic type or on igneous geological substrate tend to store SOM with high proportions of non-methoxylated aromatic constituents, which is related with comparatively low SOC levels but a high humification degree, suggested by its molecular composition indicating advanced transformation stages as regards the precursor biomass.

Capítulo 8:

Climatic variability reflected by the pyrolytic signature of soil organic matter from Mediterranean ecosystems

Este capítulo ha dado lugar a un artículo que se encuentra en proceso de revisión: Jiménez-González M.A., Álvarez A.M., Carral P., González-Pérez J.A., Almendros G. Climatic variability reflected by the pyrolytic signature of soil organic matter from Mediterranean ecosystems.

Abstract

Soil organic matter (SOM) is a major component of the biogeochemical cycle contributing to soil general properties and conservation. The progressive SOM depletion in Mediterranean ecosystems results into increased erosion and desertification. However, SOM not only plays a crucial role in soil resilience, but also represents a repository of environmental information on soil forming factors in particular climatic constraints. In this research, analytical pyrolysis (Py-GC/MS) is used to study SOM composition in 30 Spanish soils under contrasting bioclimatic scenarios as defined by the classical Q Emberger's pluviothermic quotient. Partial least squares (PLS) regression using the major pyrolysis compounds as descriptors allowed to predict ($P < 0.05$) the Q index, also suggesting the molecular proxies most responsive of the climatic variability. In addition, pyrolytic compound assemblages from soils developed under extreme climatic conditions were compared using a graphical approach based on surface density plots built from the major 193 pyrolysis compounds represented in the plane defined by their atomic ratios. The differences between the proportions of the individual pyrolysis compounds in terms of the different bioclimatic scenarios were also illustrated by a simulation of SOM molecular composition under extreme conditions of aridity or wetness. Although no cause-to-effect is inferred, the results show that the SOM molecular structure retains environmental information on the Q pluviothermic quotient, mainly reflected in the total abundances of methoxyphenols and alkylbenzene compounds. This suggest that the evolution of the SOM stabilizes in more or less advanced stages of transformation as a function of climate change.

8.1 Introduction

The progressive degradation of soils due to climatic factors in conjunction with unsuitable environmental management and agricultural practices is a global issue of current interest (Lorentz et al., 2019). In particular, the progressive degradation of soil structure, which is associated to soil organic matter (SOM) depletion with an important cementing capacity, is responsible for an increased risk of soil erosion and the substantial loss of physical and chemical fertility of soils, mainly in semiarid environments. Despite the importance of SOM in maintaining soil fertility and resilience levels (Schnitzer & Khan, 1972; Hayes & Swift, 1978; Oades, 1988; Jastrow et al., 2007), the variability on the molecular structure of the SOM encompases valuable environmental information that can be extracted from chemometric approaches applied to data supplied by spectroscopic or degradative techniques (De la Rosa et al., 2008b; Derenne & Quénéa, 2015; Angst et al., 2017).

This fact is also important, since humification and mineralization of SOM determine the balance of processes resulting in the release of CO_2 to the atmosphere, with a postulated effect on global warming (Swift, 2001). These processes are known to depend on environmental constraints i.e. average climatic conditions and their variation throughout the year. The Emberger's quotient (*Q*), calculated from annual precipitation and temperature averages (Emberger, 1932) is frequently used to estimate erosion and desertification risks, for this reason we used this index to evaluate the different scenarios in this study.

As regards analytical approaches to study the molecular structure of SOM, pyrolysis gas chromatography mass spectrometry (Py-GC/MS) is a fast and reproducible technique applicable to the direct characterization of macromolecular materials of diverse origin (Derenne & Quénéa, 2015). Py-GC/MS involves SOM thermolytic degradation into smaller fragments amenable to gas chromatography-mass spectrometry. In fact, extensive previous literature coincides in pointing out that the analyzed fragments retain the structure of the original macromolecular units (Martín et al 1972; Leinweber & Schulten, 1999). For this reason, Py-GC/MS yields a wide range of source indicator compounds for the origin of most natural, even recalcitrant, macromolecular fractions in SOM, e.g., anhydrosugars and furans from polysaccharides, methoxyphenols

from lignin, N-containing molecules from protein and chitin (Leinweber & Schulten, 1999; González-Vila et al., 2001) and condensed aromatics from black carbon (Quénéa et al., 2005a,b; De la Rosa et al., 2009; González-Pérez et al., 2014). In particular, Py-GC/MS is especially suitable for extracting information from the aliphatic fraction (mainly alkane series) of environmental samples, that conserve valuable information about the origin and transformation of the SOM (Jiménez-González et al., 2018). This technique has been recently used to monitor changes in SOM composition that can be related with different types of environmental impacts (Miralles et al., 2015; Jiménez-González et al., 2016). In this context, Py-GC/MS provides abundant chemical information on the environmental conditions in which this SOM was formed and to the direct and indirect effects on soil properties i.e. degradability, C sequestration potential, etc.

In this study the hypothesis that in Mediterranean ecosystems the molecular composition of the SOM can vary significantly as a function of aridity levels was tested using partial least squares regression (PLS) as a predictive statistical chemometric approach. In a second stage, the climate-sensitive SOM constituents were identified and selected by different approaches: in terms of the variable importance in the prediction (VIP) calculated for the individual compounds, then making quantitative comparisons between proportions of the compounds after averaging pyrochromatographic results from soils at extreme quartiles of the distribution of the values of the Emberger's index. Finally, the presumable effect of climatic changes on the molecular composition of SOM could be forecasted by simulations of the effect of variations in the Emberger's index on the proportions of the different structural components of the SOM.

8.2 Materials and Methods

8.2.1 Study area

Thirty soils from forest and cleared ecosystems from Spain were selected attending to local climatic conditions variability and classified according to the IUSS Working Group WRB system (Table 1). The sampling sites encompassed a wide range in SOC content and SOM quality (Jiménez-González et al., 2017,

2018). Topsoil samples (0–10 cm) were taken in three points and composite samples were also prepared by mixing soil material collected from three different points. The soil samples were air-dried, homogenized and sieved to fine earth (< 2 mm) before analysis.

The Emberger's index was calculated for each location according the equation:

$$Q = \frac{100 \times P}{(M^2 - m^2)}$$

where *P* = annual precipitation (mm); *M* = average of the highest temperature of the warmest month, *m* = average of the lowest temperature of the coldest month (Emberger, 1932).

8.2.2 Analytical pyrolysis

The molecular composition of the SOM was analyzed by Py-GC/MS applied to whole soil samples grinded to fine powder (< 0.01 mm) and homogenized in a planetary agate ball mill. The samples (1–2 mg) in weight were pyrolyzed for 1 min into a preheated micro-furnace at 500 °C. The instrument used consisted of a multi-shot pyrolyzer unit (Frontier Laboratories PY-2020iD) coupled to a gas chromatograph (Agilent 6890N GC) fitted with a phenylmethylsiloxane column (Agilent HP-5MS 5%). The GC oven temperature was set to 50 °C for 1 min, then increased to 100 °C at 30 °C min^{-1}; from 100 to 300 °C to a rate of 10 °C min^{-1}, and finally isothermal at 300 °C for 10 min, with a total analysis time of 32 min. The carrier gas used was helium at a flow rate of 1 cm^3 min^{-1}. The mass spectra were acquired using a quadrupole mass spectrometer (Agilent 5973N MSD) at 70 eV ionizing energy. Compound assignment was achieved via single-ion monitoring for various homologous series, via low-resolution mass spectrometry and by comparison with published and stored (NIST and Wiley libraries) data. The peak areas (total area counts) of the different compounds in the chromatograms were integrated and expressed as total abundances.

8.2.3 Partial least squares regression

The PLS regression analysis was performed using the ParLeS software (Viscarra Rossel, 2008) applied to explain the variance of Emberger's index (as dependent variable) in terms of the descriptors (independent variables) consisting of the 193 major compounds identified by Py-GC/MS (Wold et al.,

2001). To select the optimum number of factors or latent variables of each PLS model, simultaneously rejecting any possible spurious model due to overfitting, three independent criteria were considered during the cross-validation with the leave-one-out method: a) the root mean square error (RMSE), b) the Akaike's information criterion (AIC) (Akaike, 1974) and c) repeating PLS models with fully random permutation of dependent variables (models were discarded if some significant ($P < 0.05$) cross-validation function is obtained when using random Q values).

Finally, the variable importance in the projection (VIP) values were also determined for each descriptor (pyrolysis compound). This is useful to compare the weight of each compound to identify the most informative independent variables to predict the dependent environmental variable.

8.2.4 Van Krevelen plots

The compounds identified were essentially the same in all soil samples, with main differences in their relative abundances. In order to obtain a simplified and perceptual representation of the results we used an improvement of the classical van Krevelen plot (van Krevelen, 1950). This graphical-statistical approach, based on representing the compounds in a scatterdiagram in the plane defined by the O/C and H/C atomic ratios, has proven to facilitate the interpretation of complex assemblages of individual compounds released by different analytical methods. In this way, complex and abundant data can be described taking advantage of the groups formed by compounds with similar elemental composition, here corresponding to families of pyrolysis compounds of common structure (viz. alkanes, olefins, phenols...) (Kramer et al., 2004; Ikeya et al., 2015; Almendros et al., 2018). The van Krevelen method can be improved by including a third variable i.e. the accumulated abundances of individual compounds as the height, and converted into volumes by applying a two-dimensional smoothing algorithm e.g., moving average, which averages the abundance values of the next compounds in the plane into well-defined 3D peaks.

In addition of using the van Krevelen surface density plots to display the molecular of individual samples, this approach is also useful to represent the average composition of a set of samples, or to illustrate differences between sample sets (subtraction of average values from two groups of samples), as well

as to represent statistical parameters, such as VIPs, that ultimately define the importance of pyrolytic compounds in the prediction of the dependent variable, in this case *Q* index. In fact, a main objective of this study is to identify the differences between the molecular composition of SOM from soils developed under dry climates with respect to those developed under comparatively wet climates. Nevertheless VIPs values only inform on the weight of the individual compounds in the PLS model, but they are all positive, then do not straightforwardly show whether the compounds (in particular those to be considered with a major value as bioclimatic proxy) predominates in wet or dry soils. To obtain new plots displaying the above differences, the 30 soils were ordered in decreasing values of the *Q* index and two subsets were considered for comparisons, i.e., soils in the first quartile (Q1, soils developed in comparatively wet climates) vs. soils in the fourth quartile (Q4, comparatively semiarid soils). The average composition of the two subsets was calculated for the 193 independent variables and subtracted to obtain one surface density plot with positive and negative values, illustrating compounds prevailing in soils formed under wet or arid climatic conditions, respectively.

To confirm whether the differences between the average values of the compounds in the two soil subsets under contrasted climatic conditions were significant, a Student's *t* was calculated comparing the abundances for each of the compounds in soils in the two quartiles. The calculation of the Student's *t*, were carried out with spreadsheet facilities of Microsoft Excel, i.e., the T.TEST function (array1,array2,tails,type), with arguments to declare 2 tails and two-sample unequal variance (heteroscedastic). The Student’s *t* values were also represented in van Krevelen plots, in this case as a transparent contour diagram, which is superimposed on the above density map representing the differences between the abundances of the compounds in soils formed under contrasted climatic conditions.

To finish with this exploratory analysis on the variability of the molecular composition of the SOM in terms of the *Q* index, a final simulation was carried out to predict the composition of the SOM for extreme climatic scenarios of high or low aridity, respectively. This was done by extrapolating the relative abundance of each compound for the desired values of the Emberger index, in two pyrochromatographic compound arrays corresponding to virtual soils with a

high *Q* value of 400 and soils with a low *Q* value of 5. The forecasted values would illustrate the hypothetical effect of climatic changes on SOM composition and would allow discussing its different quality in terms of resilience or maturity.

The Excel's function FORECAST which calculates or predicts a future value by using existing values, was used to extrapolate the relative abundances of each of the compounds in virtual soils developed under *Q* values outside the limits registered in this study.

8.3 Results

8.3.1 General characteristics of the soils and Py-GC/MS data

The Emberger's pluviothermic *Q* index in the different areas showed the desired large variability (Table 2), the highest value in the soil series being 278 (humid) and the lowest 38 (semiarid).

Up to 193 different pyrolysis compounds were identified using Py-GC/MS of whole soil samples and grouped in different families (Table 12). The main molecules identified corresponded to aliphatic series: *n*-alkanes, olefins, fatty acids and aromatic compounds (alkylbenzenes, methoxyphenols, alkylphenols and polycyclic hydrocarbons). In addition, a series of carbohydrate-derived constituents (methylfuran, furaldehydes, maltol and levoglucosan) and N-compounds (pyrroles, pyridines and indoles), from peptides and proteins, were also identified. The same pattern of lignin-derived units was found in all samples, consisting of different types of methoxyphenols, i.e., both guaiacyl (G)-lignin units (guaiacol and its methyl-, ethyl-, vinyl-, propenyl- and aceto- derivatives) and their corresponding syringyl (S) counterparts (Sáiz-Jiménez & de Leeuw, 1986). In some samples the moderate abundances of lignin-derived pyrolysis compounds is accompanied by high yields of methoxyl-lacking aromatic compounds (alkylbenzenes, alkylphenols and alkylnaphthalenes). The latter considered more characteristic of SOM at comparatively advanced stages of transformation, at least when maturity is operatively defined as the extent to which plant biomass constituents such as lignin are transformed into

macromolecular materials of chaotic structure referred to as humic substances (Almendros et al., 2018).

8.3.2 Partial least squares regression

The cross-validation plots for the PLS models are shown in Fig. 8.1. A good prediction of *Q* was obtained exclusively using the information provided by the normalized abundances of the different compounds identified by Py-GC/MS, and 6 latent variables. This was checked repeating the process using randomized values for the *Q* index, which resulted in a non-significant cross-validation function (observed vs predicted), confirming that the model was not spurious or overfitted.

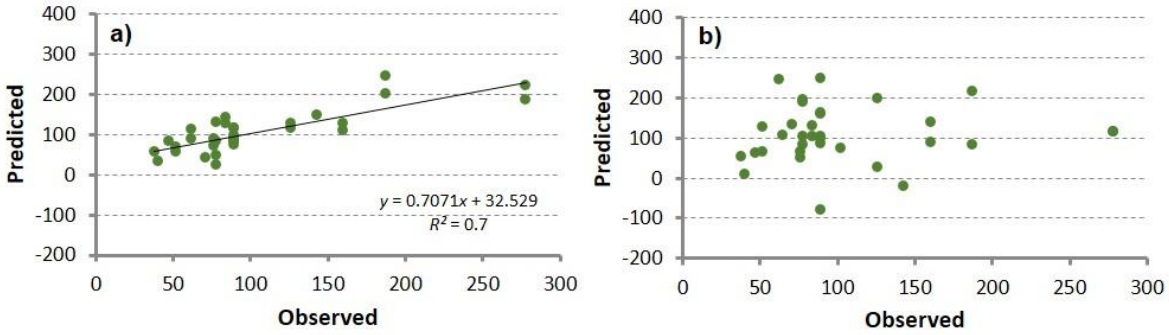


Figure 8.1. Observed versus predicted values of Emberger's index (*Q)* calculated by partial least squares (PLS) regression using pyrolytic compounds from whole soil samples as descriptors (a) compared with the cross-validation plot using the fully random permutation of *Q* index values (b).

The compounds with the highest VIP values calculated in the PLS model to forecast *Q* are shown in Fig. 8.2. The highest values were found in the region for aliphatic compounds (alkane and olefin regions) although regions for alkylaromatic compounds, phenols and methoxyphenols also showed comparatively high values.

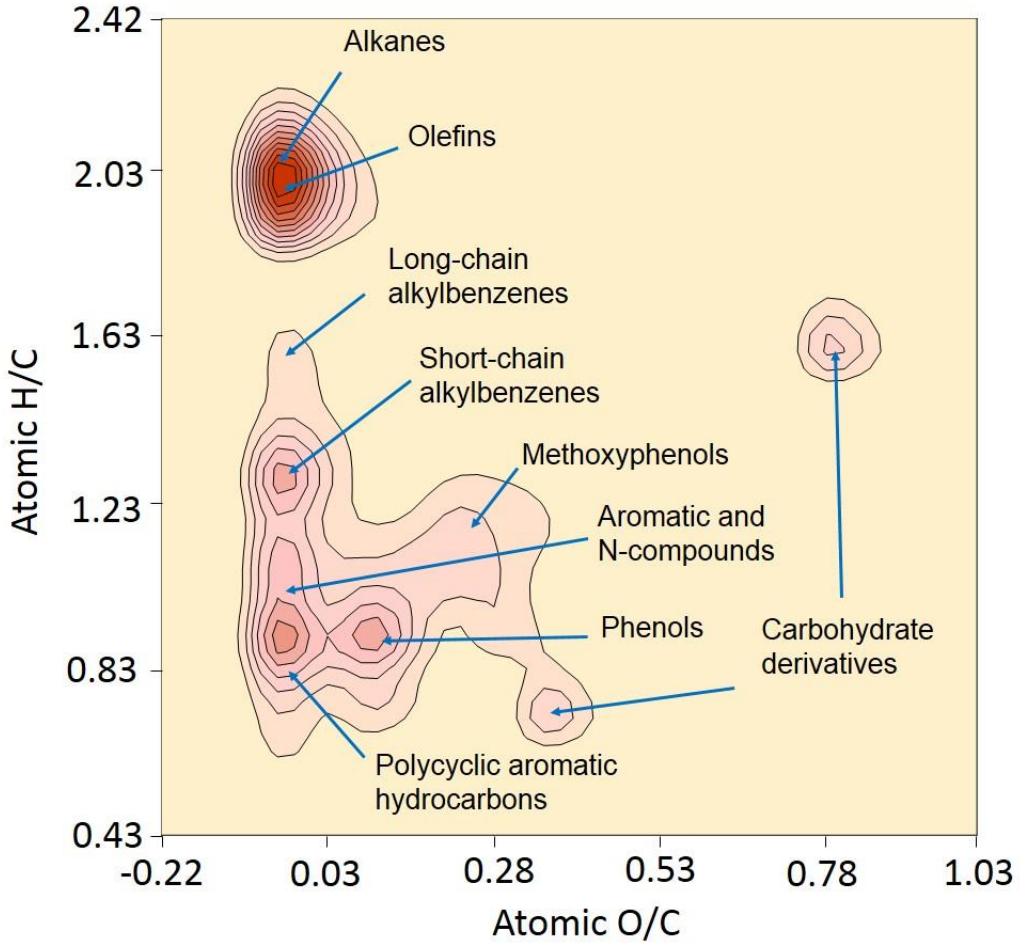


Figure 8.2. Surface density map showing values of variable importance for prediction (VIP) of different types of compounds used as descriptors in the partial least squares (PLS) models to predict the Emberger's climatic index (*Q)* from pyrolytic data of whole soil samples

8.3.3 Van Krevelen surface density plots

The van Krevelen diagrams in Fig. 8.3 show the average abundances of pyrolytic compounds from soils under contrasted values of *Q* index. For rapid identification of the major changes in the pyrolytic signature, the surface obtained by subtracting the compound abundances in the above sets is also shown. In these plots it is possible to observe a comparatively higher concentration of alkylaromatic and polycyclic aromatic compounds in soils developed in areas with high *Q* index (humid conditions: Q1). Conversely, the composition of the SOM in soils developed on low *Q* index (semiarid climate) includes comparatively higher content of methoxyphenols. Table 13 shows the

compounds whose proportion shows a major and significant difference between soils under different climate defined by the *Q* index.

Concerning the representation of pyrochromatographic signatures for two extrapolated situations, one for *Q* values higher than those of the soil with the uppermost value (*Q* = 400) and another one with a value below the lowest value (*Q* = 5), and shown in Fig. 8.4. It was observed that in particular the extent of the methoxyphenols decrease with the Emberger's index, whereas aliphatic constituents such as carbohydrate-derived, long-chain alkylbenzenes and in general alkyl compounds tend to accumulate in SOM of soils developed under humid conditions.

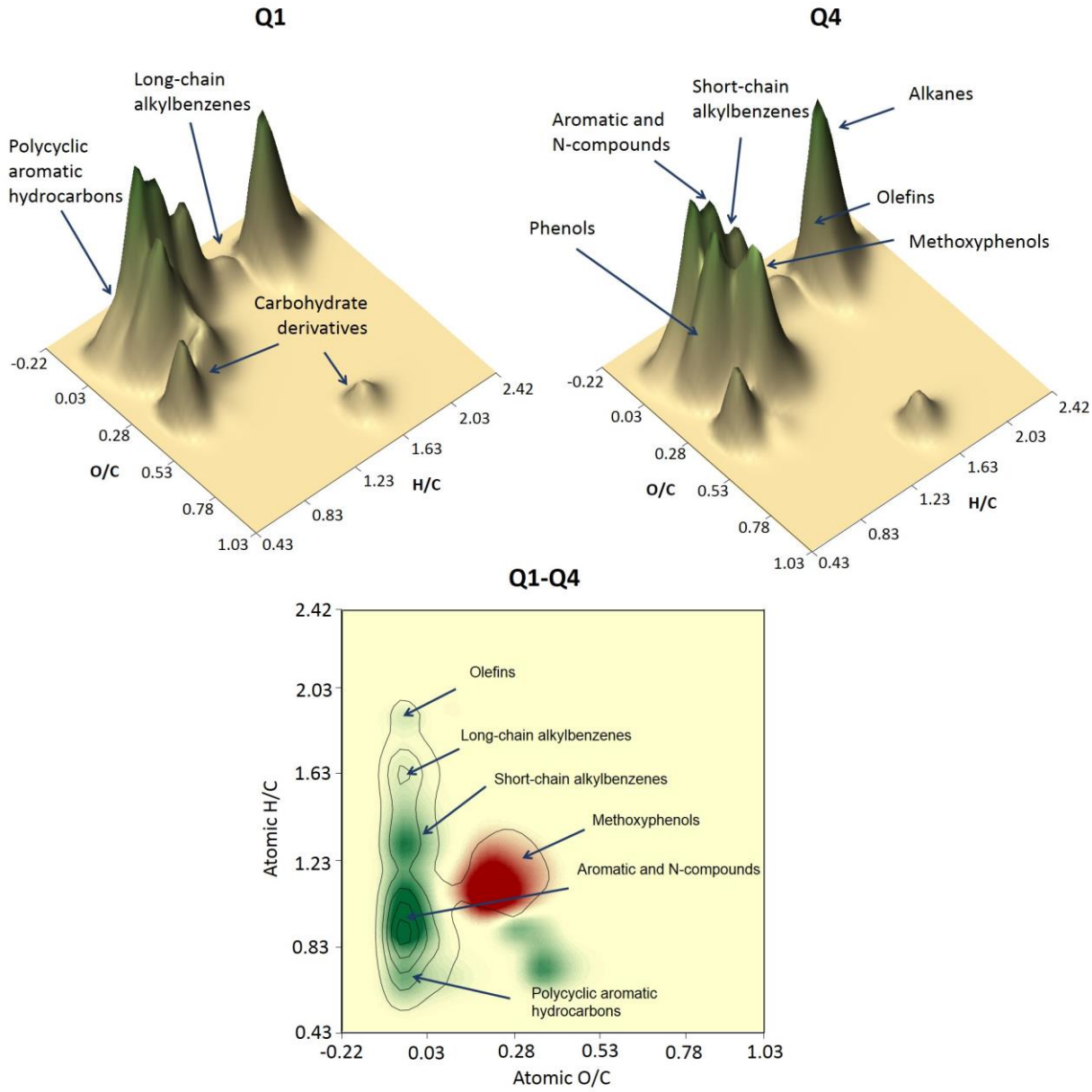


Figure 8.3. Surface density maps displaying cumulative abundances of soil pyrolysis compounds represented in the space defined by their H/C and O/C atomic ratios in a van Krevelen diagram. Plots Q1 and Q4 show the average composition of pyrochromatograms from soil samples in the quartile with high (Q1) or low (Q4) values of the Emberger's index, respectively. The subtraction values between quartiles (Q1-Q4) are represented as a 2D van Krevelen diagram, where positive values are indicated in green coluor (comparatively wet climates), and negative values in red (comparatively arid climates). The superimposed contour diagram shows the significance levels of the differences as Student's *t* values between proportions of each compound in the two soil sets, $P < 0.1$ in the external contour).

8.4 Discussion

The significant ($P < 0.05$) prediction model obtained for *Q* index using the 193 major pyrolysis compounds analyzed by Py-GC/MS as descriptors indicated that a relationship exists between the chemical composition of the SOM and the direct or indirect effect of climatic constraints in the ecosystem properties, reflected in the soil biogeochemical performance. The VIP values showed that the proportions of pyrolytic alkanes and olefins play an important role in the prediction model. This fact can be related with the quantitative information provided by the alkane composition about the relative contribution of microbial metabolism with respect to the accumulation of products inherited from vascular plants (Jiménez-González et al., 2018; Simoneit & Mazurek, 1982). In fact, despite the subsequent data analyses will show that the total proportion of aliphatic products is not substantially modified according to *Q* index, it is remarkable that even small quantitative variations in alkanes supply large information to predict the above climatic index.

In particular Table 14 shows the compounds with the highest values of VIPs from the PLS model to predict the *Q* index from pyrolytic data, i.e., those providing more information regarding the different molecular composition of the SOM in sites with different pluviothermic index. This list of compounds with large VIP values includes both short-chain alkanes (e.g., < 23 C) and comparatively long-chain alkanes. Among the proportions of these compounds, a quantitative balance is established, which depends on the origin of the SOM either microbial or from vascular plants (Jiménez-González et al., 2018; Simoneit & Mazurek, 1982). In particular, pyrolysis preserves valuable information in the form of early thermoevaporation products such as alkanes, that retain carbon preference number and relative chain length indices that inform about their origin (Jiménez-González et al., 2018).

Figure 8.2 also shows that a large part of the differences between the climatic types are reflected by the methoxyphenols, typical products derived from lignin, and that inform on the extent to which plant biomass remains non-decomposed or selectively preserved in different types of soils, in this case with proportions increasing towards arid climates.

As a difference to the VIPs, the subtraction plots between average data in extreme quartiles (Fig. 8.3) illustrates, in a semiquantitative way, the compounds characteristic for the different levels of the Emberger's index which, in some way, is also informing on potential risks of soil degradation and desertification. The higher proportions of non-methoxylated aromatic compounds indicated a predominance of comparatively transformed SOM. This is interpreted as the result of enhanced microbial activity with increasing wetness. Conversely, SOM formed on semiarid conditions would display high levels of plant-inherited, lignin-derived compounds (Sáiz-Jiménez & de Leeuw, 1986; Hayes et al., 1989; Zech et al., 1989; Stevenson, 1994; Almendros et al., 2018).

This fact could be associated to comparatively lower microbial activity in dry conditions. The above differences are more easily identified in the extrapolated situations, where the remarkable increase in methoxyphenols can be observed in an ideal situation of extremely dry climatic conditions observed in Fig. 8.4.

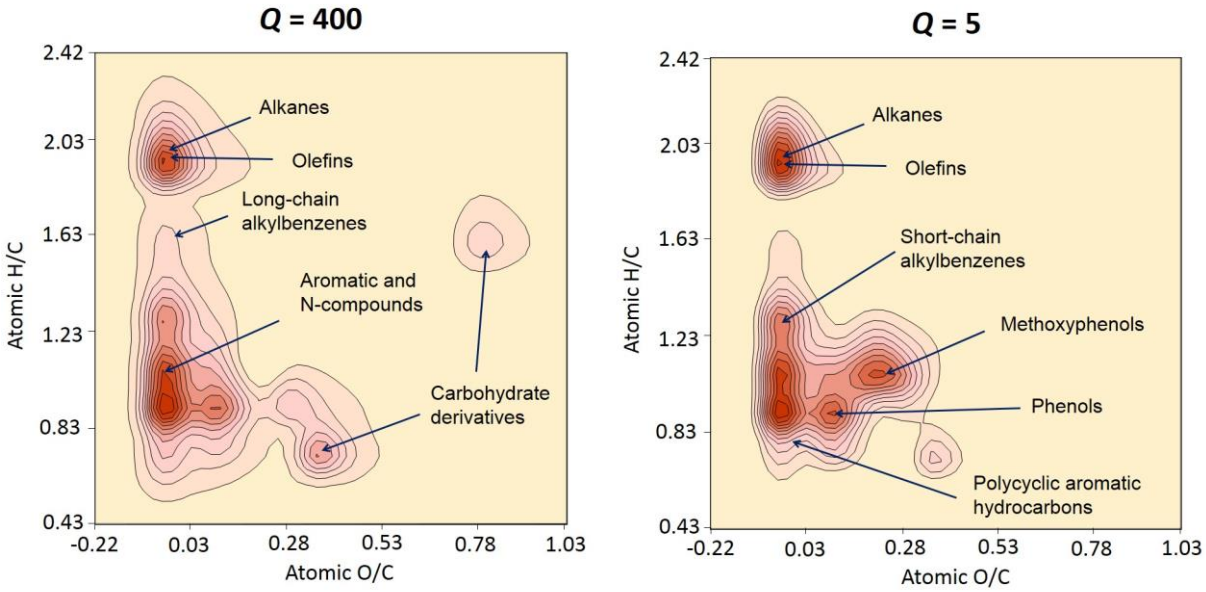


Figure 8.4. Comparison of van Krevelen surface density maps illustrating the composition of soil pyrolysis compounds extrapolated for virtual scenarios with high Emberger's index (*Q* = 400, humid), and extreme low *Q* index (*Q* = 5, arid).

When the quantitative differences between the proportions of pyrolysis compounds from soils developed under contrasted values of the Emberger index is described for individual compounds it would be possible to identify those

which would present a higher value as bioclimatic molecular proxies. Table 13 show a list of the compounds the normalized concentration of which varies more than 50% between the groups of samples (humid, arid) studied.

According to the conclusions suggested by the density surfaces, the nature of the compounds in Table 14 shows the importance of differentiating between aromaticity associated to lignin components with respect to that depending on condensed and/or alkyl-substituted aromatic structures. In fact, more than the total aromaticity, the ratio between these two types of aromatic compounds would reflect to a greater extent the impact of environmental factors on the soils (Tinoco et al., 2015).

8.5 Conclusions

The Py-GC/MS present a large potential not only as fingerprinting technique, but also to identify the origin and transformation of the SOM and to monitor environmental changes reflected in its molecular composition. In this case, it could be demonstrated that a reduced set of the 193 major pyrolysis compounds had an interesting value as bioclimatic proxies, since their proportions vary significatively paralleling the different levels of aridity or humidity in the ecosystems where the SOM was accumulated. Soils developed under semiarid conditions tend to accumulate increased proportions of methoxyphenol structures suggesting the presence of altered lignins. Conversely, condensed and alkyl-substituted aromatic compounds considered more characteristic of resilient SOM at advanced stages of transformation, were found in soils developed under comparatively humid conditions. From the quantitative point of view, long-chain aliphatic constituents of the SOM were especially stable in terms of changes in climatic conditions.

Judging from the behavior of the 30 Spanish soils studied—in general under continental Mediterranean climate—it seems that a hypothetical climate change towards greater aridity would result in a lower quality of the SOM, which would accumulate structural constituents inherited from lignocelluloses at intermediate stages of transformation.

Conclusiones Generales

En esta Tesis Doctoral se ha analizado un total de 35 suelos de España con una gran variabilidad en sus propiedades físico-químicas, con especial referencia a su diferente contenido en carbono. En este sentido, la relación entre la cantidad de la SOM y su composición química sigue siendo objeto de estudio en la actualidad. El objetivo principal de este trabajo se ha centrado en el conocimiento de los factores ambientales que podrían estar ejerciendo una función significativa en la estabilidad de la SOM y por consiguiente en la conservación de la calidad de los suelos. Para este objetivo global se marcaron varios objetivos específicos que se han estudiado en cada uno de los capítulos de esta Tesis Doctoral.

- En el primer objetivo se planteó estudiar la composición de los alcanos, hidrocarburos saturados, presentes en la composición de la SOM. En el capítulo 3 se describió como la diversidad de estos compuestos, reflejada por el índice de Shannon, se encuentra correlacionada con el contenido de carbono de los suelos, sugiriendo que la complejidad química de la SOM favorece su estabilidad. En particular, la correlación con el contenido en SOM es más significativa en el caso de los alcanos de cadena larga (C_{24}–C_{31}), mientras que los alcanos de cadena media actúan como indicadores de la su calidad. Finamente, los alcanos de cadena más corta (C_9–C_{18}) no muestran correlación significativa con las características del suelo, pudiendo ser debido a que suelen provenir de la ruptura de cadena en las condiciones de pirólisis analítica utilizada para su análisis.

- Como segundo objetivo, se planteó el estudiar la medida en que los precursores de la SOM podrían explicar su calidad y contenido. Para ello, se seleccionaron los 12 principales metoxifenoles obtenidos por pirólisis analítica de la SOM. La diversidad de estos compuestos, expresada como el índice de Shannon, mostró una relación con la concentración de SOC, lo que sugiere que la acumulación de formas estables de C en el suelo depende de la naturaleza de la biomasa vegetal y de la complejidad adquirida durante su transformación en el suelo. Se encontró que la mayor diversidad en metoxifenoles concurre en los suelos con mayor contenido en SOM. Por otro lado, se comprobó como esta distribución de metoxifenoles en el suelo depende significativamente del tipo de vegetación predominante (gimnospermas o angiospermas).

- La parte del estudio que se llevó a cabo sobre la fracción de HAs aislados del suelo, y que ha sido descrita en los capítulos 5 y 6, ayuda a interpretar la información química que puede extraerse del análisis de su composición, y relacionarla con el contenido de SOC. Los resultados mostraron una clara relación entre su composición molecular y el contenido de SOC en el suelo a través de modelos de PLS. Se concluyó que existía una prevalencia de estructuras de lignina en la composición de los HAs aislados de suelos con alto contenido en SOC, mientras que en suelos de bajo SOC se acumulaban HAs con abundancia en grupos carbonilo, amida y estructuras aromáticas. Una tendencia inversa se observa con la E4 (densidad óptica de las soluciones de HA, indicadora de su madurez), de tal forma que mayores valores de calidad van ligados a grupos aromáticos, carboxilo y amidas mientras que la presencia de lignina va ligada a bajos valores de E4. En el caso de la composición molecular de los HAs estudiada por FTICR-MS, se observó igualmente una relación a través de PLS entre las fórmulas moleculares encontradas en todos los HA y el contenido de SOC. Suelos con altos niveles de SOC muestran una acumulación de lípidos insaturados y ligninas en los HA, mientras que a bajos niveles de SOC predominan los lípidos saturados. Un hecho relevante a destacar es que estos resultados muestran que los HAs conservan valiosa información incluso después de la extracción en medio alcalino con pH elevado.

- Los factores formadores del suelo presentan gran influencia en la composición de la SOM y en sus niveles de carbono. Dichos factores influyen en los procesos de transformación de la SOM, siendo el más significativo el efecto de la vegetación, seguido por el clima y en menor medida el sustrato geológico. Se encontró que los suelos desarrollados bajo vegetación de gimnospermas o sobre sustrato sedimentario, mostraron los niveles más elevados de carbono, pero con una SOM con características típicas de la lignina, lo que refleja escasa transformación. En cambio, suelos con bajo contenido en SOC como son los desarrollados en clima Cfb o sustrato ígneo, muestran alto contenido en componentes aromáticos no metoxilados, indicativos de un avanzado estado de humificación.

- Como último objetivo se planteó examinar si la composición de la materia orgánica podría reflejar la variación del estado del suelo de acuerdo con cambios en parámetros climáticos, y a su vez con el avance de la desertificación. Se ha comprobado que la composición de la SOM varía significativamente en función de indicadores climáticos tales como el índice de Emberger. Suelos bajo condiciones semiáridas tienden a acumular metoxifenoles, derivados de la lignina. Por otra parte, condiciones más húmedas de los suelos favorecen la presencia de estructuras aromáticas alquil sustituidas o condensadas, indicadoras de un avanzado estado de transformación. Esto sugiere que en un futuro escenario de cambio climático con avance de la desertificación, la calidad de la SOM se vería afectada, acumulándose mayor proporción de constituyentes derivados de la lignina en fases tempranas de transformación.

Todas estas conclusiones muestran una clara relación entre la composición molecular o calidad de la SOM y su contenido total en el suelo. Parece existir una tendencia bastante constante a que suelos que acumulan mayores niveles de carbono lo hagan a expensas de la preservación de una SOM menos humificada, mas lábil, mientras que suelos con poco carbono suelen presentar una SOM más recalcitrante, incluyendo constituyentes aromáticos y condensados. Esta variabilidad composicional de la SOM acumulada bajo distintos escenarios puede explicarse a partir de la influencia de los factores formadores del suelo, cuyo impacto queda reflejado en la composición molecular de la SOM, permitiendo proponer modelos de predicción de los cambios esperados en su estructura y propiedades por efecto del avance de la desertificación o incremento de la aridez en los suelos mediterráneos.

General conclusions

In this PhD study, 35 soils from Spain with large variability in their physical-chemical properties, in particular its C content, have been analyzed. In this context, the relationship between the amount of SOM and its chemical composition has been the subject of special attention. The main objective of this work is focused on the knowledge of the environmental factors that could be playing a significant function in the stability of the SOM and with regard to the conservation of soil quality. For this global objective, several specific objectives have been developed in each of the titles of this PhD study:

- The first objective was to study the composition of alkanes, saturated hydrocarbons, which are present in the composition of the SOM. In chapter 3, it was described how the diversity of these compounds, reflected by the Shannon index, is correlated with the carbon content, suggesting that the chemical complexity of SOM favors its stability. In particular, the correlation with the SOM content is more significant in the case of long-chain alkanes (C_{24}–C_{31}), while medium-chain alkanes are indicators of their quality. Finally, the shorter chain alkanes (C_9–C_{18}) do not show a significant correlation with soil characteristics, which is possibly due to they usually come from the chain break in the conditions of analytical pyrolysis used for analysis.

- In the second objective, it was proposed to study the extent to which SOM precursors could explain its quality and content. For this purpose, the 12 main methoxyphenols released by analytical pyrolysis of the SOM were selected. The diversity of these compounds, expressed as the Shannon index, showed a relationship with the concentration of SOC, which suggests that the accumulation of stable forms of C in the soil depends on the nature of the plant biomass and its complexity acquired during Its transformation in the soil. It was found that the greatest diversity in methoxyphenols occurs in soils with higher SOM content. On the other hand, it was observed that this distribution of methoxyphenols depends significantly on the type of predominant vegetation (gymnosperms or angiosperms).

- The part focused on the study of HAs, corresponding to chapters 5 and 6, helps to interpret the chemical information of them and its bearing with the SOC content. The results showed a relationship between chemical composition of HAs and the SOC content using PLS models. It was

concluded that there was a prevalence of lignin structures in the composition of HA isolated from soils with a high SOC content, while in soils with low SOC, HAs consist mainly of aromatic, carboxyl and amide structures. An inverse trend is observed with E4, so that higher SOM quality is linked to carboxyl, aromatic and amide groups, whereas predominance of lignin is linked to low E4 values. In the case of the molecular composition of HAs studied by FTICR-MS, a relationship was also observed between the compounds found in all HAs and the SOC content. Soils with high levels of SOC show an accumulation of unsaturated lipids and lignins in the composition of HA, while at low levels of SOC saturated lipids predominate. A relevant fact to mention is that the whole above results show that HA structure contains valuable biogeochemical information after extraction with alkaline medium.

- Soil forming factors have a great influence on the composition of the SOM and soil carbon levels. These factors influence the transformation processes of the SOM, the effect of vegetation is the most significant, followed by the climate and to a lesser extent the geological substrate. It was found that soils developed under gymnosperm vegetation or on sedimentary substrate, showed high levels of carbon, but with the SOM showing typical characteristics of lignin, reflecting poor transformation. On the other hand, soils with low SOC content, such as those developed in Cfb climate or igneous substrate, show high content of non-methoxylated aromatic components, indicative of an advanced state of humification.

- In the final objective, it was examined whether the composition of the organic matter could reflect the variation in climatic parameters, and in turn the progress of desertification. It has been observed that the composition of the SOM varies significantly according to climatic indicators such as the Emberger index. Soils under semi-arid conditions tend to accumulate methoxyphenols, indicating a recent origin from lignin. On the other hand, soil moisture favors the presence of alkyl-substituted or condensed aromatic structures, indicative of an advanced state of transformation. This suggests that in a future scenario of climate change with progressive desertification, the quality of SOM would be affected, accumulating a greater proportion of constituents derived from lignin in the early stages of transformation.

All these conclusions show a clear relationship between the molecular composition or quality of SOM and its total content in soil. There seems to be a constant tendency for soils that accumulate higher levels of carbon at the expense of preserving a less humified, more labile SOM, while soils with low carbon content tend to have a more recalcitrant SOM, including aromatic and condensed constituents. This compositional variability of the SOM accumulated under different scenarios can be explained from the influence of the soil forming factors, whose impact is reflected in the molecular composition of the SOM, allowing to propose models of prediction of the expected changes in its structure and properties as a result of the progress of desertification or increased aridity in Mediterranean soils.

Bibliografía

Abdulla H.A.N., Minor E.C., Dias R.F., Hatcher P.H., 2010. Changes in the compound classes of dissolved organic matter along an estuarine transect: a study using FTIR and ^{13}C NMR. *Geochimica et Cosmochimica Acta* 74:3815–3838.

Aitchison J., 1986. The statistical analysis of compositional data. In: Monographs on Statistics and Applied Probability. London: Chapman & Hall.

Akaike H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–723.

Almendros G., Sanz J., González-Vila F.J., Martín F., 1991. Evidence for a polyalkyl nature of soil humin. *Naturwissenschaften* 78:359–362.

Almendros G., González-Vila F.J., Martín F., Fründ R., Lüdemann H.-D., 1992. Solid state NMR studies of fire-induced changes in the structure of humic substances. *Science of the Total Environment* 117–118:63–74.

Almendros G., Dorado J., 1999. Molecular characteristics related to the biodegradability of humic acid preparations. *European Journal of Soil Sceince* 50:227–236.

Almendros G., 2008. Humic substances. En: Chesworth W. (Ed.). Encyclopedia of Soil Science. Springer. Dordretch, pp. 97–99.

Almendros G., Tinoco P., De la Rosa J.M., Knicker H., González-Pérez J.A., González-Vila F.J., 2016. Selective effects of forest fires on the structural domains of soil humic acids as shown by dipolar dephasing ^{13}C NMR and graphical-statistical analysis of pyrolysis compounds. *Journal of Soils and Sediments* 18:1303–1313.

Almendros G., Hernández Z., Sanz J., Rodríguez-Sánchez S., Jiménez-González M.A., González-Pérez J.A., 2018. Graphical statistical approach to soil organic matter resilience using analytical pyrolysis data. *Journal of Chromatography A* 1533:164–173.

Amblès A., Jacquesy J.C., Jambu P., Joffre J., Maggi-Churin R., 1991. Polar lipid fraction in soil: a kerogen-like matter. *Organic Geochemistry* 17:341–349.

Angst Š., Mueller C.W., Cajthaml T., Angst G., Lhotáková Z., Bartuska M., Spaldoňová A., Frouz J., 2017. Stabilization of soil organic matter by

earthworms is connected with physical protection rather than with chemical changes of organic matter. *Geoderma* 289:29−35.

Aoyama T., Terasawa S., Sudo K., Shiori T., 1984. New Methods and Reagents in Organic Synthesis. 46. Trimethylsilyldiazomethane: A convenient reagent for the *O*-methylation of phenols and enols. *Chemical and Pharmaceutical Bulletin* 32:3759–3760.

Baldock J.A., Oades J.M., Waters A.G., Peng X., Vassallo A.M., Wilson M.A., 1992. Aspects of the chemical structure of soil organic materials as revealed by solid-state ^{13}C NMR spectroscopy. *Biogeochemistry* 16:1–42.

Banerjee S., Kirkby C.A., Schmutter D., Bissett A., Kirkegaard J.A., Richardson A.E., 2016. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil Biology and Biochemistry* 97:188–198.

Barré P., Quénéa K., Vidal A., Cécillon L., Christensen B.T., Kätterer T., Macdonald A., Petit L., Plante A.F., van Oort F., Chenu C., 2018. Microbial and plant-derived compounds both contribute to persistent soil organic carbon in temperate soils. *Biogeochemistry* 140:81–92.

Batjes N.H., 1996. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science* 47:151–163.

Baveye P.C., Wander M., 2019. The (bio)chemistry of soil humus and humic substances: why is the “new view” still considered novel after more than 80 years?. *Frontiers in Environmental Science* 7:27.

Blake G.R., Hartge K.H., 1986. Particle density. In: Klute (Ed.). Methods of Soil Analysis: Part 1—Physical and mineralogical methods. American Society of Agronomy and Soil Science Society of America, Madison, pp. 363–375.

Bockheim J.G., Hartemink A.E., 2017. The Soils of Wisconsin. Springer. Madison.

Boerjan W., Ralph J., Baucher M., 2003. Lignin biosynthesis. *Annual Review of Plant Biology* 54:519–546.

Bonawitz N.D., Chapple C., 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annual Review of Genetics* 44:337–363.

Bouyoucos G.J., 1927. The hydrometer as a new method for the mechanical analysis of soils. *Soil Science* 23:343–353.

Bouyoucos G.J., 1962. Hydrometer method improved for making particle size analysis of soils*. Agronomy Journal* 54:464–465.

Brady N.C., 1990. The nature and properties of soils. 10th Edition, Macmillan Publishing Company, Cranbury.

Buurman P., Peterse F., Almendros G., 2007. Soil organic matter chemistry in allophanic soils: a pyrolysis-GC/MS study of a Costa Rican Andosol catena. *Soil Science* 58:1330–1347.

Cécillon L., Certini G., Lange H., Forte C., Strand L.T., 2012. Spectral fingerprinting of soil organic matter composition. *Organic Geochemistry* 46:127–136.

Chechevatov A.I., Miroshnichenko Y.S., Myasoyedova T.N., Popov Y.V., Yalovega G.E., 2017. Investigations of the capability to heavy metals adsorption humic acids: correlation between structure and absorption properties. In: Parinov I., Chang S.H., Jani M. (Eds) Advanced Materials. Springer Proceedings in Physics, vol 193. Springer, Cham.

Chen Y., Huang Y., Sun W., 2017. Using organic matter and pH to estimate the bulk density of afforested/reforested soils in northwest and northeast China. *Pedosphere* 27:890–900.

Dabin B., 1976. Méthode d'extraction et de fractionnement des matières humiques du sol: application à quelques études pédologiques et agronomiques dans les sols tropicaux. Cah. ORSTOM Série Pédologie 14:287–297.

Dabin B., 1971. Étude d'une méthode d'extraction de la matière humique du sol. *Science du Sol* 1:47–63.

D'Acqui L.P., Pucci A., Janik L.J., 2010. Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science* 61:865–876.

De la Rosa J.M., González-Pérez J.A., González-Vázquez R., Knicker H., López-Capel E., Manning D.A.C., González-Vila F.J., 2008a. Use of pyrolysis/GC–MS combined with thermal analysis to monitor C and N changes

in soil organic matter from Mediterranean fire affected forest. *Catena* 74:296–303.

De la Rosa J.M., González-Pérez J.A., Hatcher P.G., Knicker H., González-Vila F.J., 2008b. Determination of refractory organic matter in marine sediments by chemical oxidation, analytical pyrolysis and solid-state ^{13}C nuclear magnetic resonance spectroscopy. *European Journal of Soil Science* 59:430–438.

De la Rosa J.M., González-Vila F.J., López-Capel E., Knicker H., González-Pérez J.A., 2009. Structural properties of non-combustion-derived refractory organic matter which interfere with BC quantification. *Journal of Analytical and Applied Pyrolysis* 85:399−407.

De la Rosa J.M., Jiménez-Morillo N.T., González-Pérez J.A., Almendros G., Vieira D., Knicker H.E., Keizer J., 2019. Mulching-induced preservation of soil organic matter quality in a burnt eucalypt plantation in central Portugal. *Journal of Environmental Management* 231:1135–1144.

Derenne S., Quénéa K., 2015. Analytical pyrolysis as a tool to probe soil organic matter. *Journal of Analytical and Applied Pyrolysis* 111:108−120.

Di X., Xiao B., Dong H., Wang S., 2019. Implication of different humic acid fractions in soils under karst rocky desertification. *Catena* 174:308–315.

DiDonato N., Chen H., Waggoner D., Hatcher, P.G., 2016. Potential origin and formation for molecular components of humic acids in soils. *Geochimica et Cosmochimica Acta* 178:210–222.

Dillon M.L., 2011. Explaining soil organic carbon sequestration in an urban ecosystem, 96th ESA Annual Meeting. COS 193-7, Austin, Texas.

Dinel H., Lévesque M., Mehuys G.R., 1991. Effects of long-chain aliphatic compounds on the aggregate stability of a lacustrine silty clay. *Soil Science* 151:228–239.

Duan Y., He Y., 2011. Distribution and isotopic composition of *n*-alkanes from grass, reed and tree leaves along a latitudinal gradient in China. *Geochemical Journal* 45:199–207.

Duchaufour P., 1970. Précis de Pédologie. Masson & Cie. Paris.

Duchaufour P., Jacquin F., 1975. Comparaison des procesus d'humification dans les principaux types d'humus forestiers. *Science du Sol* 1:29–36.

Dungait J.A.J., Hopkins D.W., Gregory A.S., Whitmore A.P., 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change Biology* 18:1781–1796.

Eglinton G., Hamilton R.J., 1967. Leaf epicuticular waxes. *Science* 156:1322–1335.

Eglinton G., Logan G.A., 1991. Molecular preservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 333:315–328.

Emberger L., 1932. Sur une formule climatique et ses applications en botanique. *La Météorologie* 92:1–10.

Faria S.R., De la Rosa J.M., Knicker H., González-Pérez J.A., Villaverde, J., Keizer, J.J., 2015. Wildfire-induced alterations of topsoil organic matter and their recovery in Mediterranean eucalypt stands detected with biogeochemical markers. *European Journal of Soil Science* 66:699–713.

Farmer V.C., Morrison R.I., 1960. Chemical and infrared studies of Phragmites peat and its humic acid. *Scientific Proceedings of the Royal Dublin Society* 1:85–104.

Fengel D., Wegener G., 1984. Wood: Chemistry, Ultrastructure, Reactions. Walter de Gruyter, Berlin and New York.

Fernández-Getino A.P., Hernández Z., Piedra Buena A., Almendros G., 2013. Exploratory analysis of the structural variability of forest soil humic acids based on multivariate processing of infrared spectral data. *European Journal of Soil Science* 64:66–79.

Ganuza A., Almendros G., 2003. Organic carbon storage in soils of the Basque Country (Spain): the effect of climate, vegetation type and edaphic variables. *Biology and Fertility of Soils* 37:154–162.

Glazer A.W., Nikaido H., 1995. Microbial Biotechnology: Fundamentals of Applied Microbiology. San Francisco, W.H. Freeman, p. 340.

Guitián F., Carballas T., 1976. Técnicas de Análisis de Suelos. Ed. Pico Sacro, Santiago de Compostela, España.

Gocke M., Kuzyakov Y., Wiesenberg G.L.B., 2013. Differentiation of plant derived organic matter in soil, loess and rhizoliths based on *n*-alkane molecular proxies. *Biogeochemistry* 112:23–40.

González-Pérez J.A., González-Vila F.J., González-Vázquez R., Arias M.E., Rodríguez J., Knicker H., 2008. Use of multiple biogeochemical parameters to monitor the recovery of soil after forest fires. *Organic Geochemistry* 39:940-944.

González-Pérez J.A., Almendros G., De la Rosa J.M., González-Vila F.J., 2014. Appraisal of polycyclic aromatic hydrocarbons (PAHs) in environmental matrices by analytical pyrolysis (Py–GC/MS). *Journal of Analytical and Applied Pyrolysis* 109:1−8.

González-Vila F.J., Lüdemann H-D., Martín F., 1983. ^{13}C-NMR structural features of soil humic acids and their methylated, hydrolyzed and extracted derivatives. *Geoderma* 31:3–15.

González-Vila F.J., Tinoco P., Almendros G., Martin F., 2001. Pyrolysis−GC−MS Analysis of the formation and degradation stages of charred residues from lignocellulosic biomass. *Journal of Agricultural and Food Chemistry* 49:1128–1131.

Hayes M.H.B., Swift R.S., 1978. The Chemistry of Soil Organic Colloids D.J. Greenland, M.H.B. Hayes (Eds.), The Chemistry of Soil Constituents, Wiley, Chichester, pp. 179–320.

Hayes M.H.B., MacCarthy P., Malcom R.L., Swift R.S., 1989. Humic substances II. Search of Structure. Wiley, New York.

Hayes M.H.B., 1991. Advances in Soil Organic Matter Research: The Impact on Agriculture and the Environment. Woodhead Publishing, Sawston.

Hernández Z., Almendros G., 2012. Biogeochemical factors related with organic matter degradation and C storage in agricultural volcanic ash soils. *Soil Biology and Biochemistry* 44:130–142.

Hernández Z., Almendros G., Carral P., Álvarez A., Knicker H., Pérez Trujillo J.P., 2012. Influence of non-crystalline minerals in the total amount, resilience

and molecular composition of the organic matter in volcanic ash soils (Tenerife Island, Spain). *European Journal of Soil Science* 63:603–615.

Hernández Z., Almendros G., Álvarez A., Figueiredo T., Carral P., 2019. Soil carbon stabilization pathways as reflected by the pyrolytic signature of humic acid in agricultural volcanic soils. *Journal of Analytical and Applied Pyrolysis* 137:14–28.

Ikeya K., Sleighter R.L., Hatcher P.G., Watanabe A., 2015. Characterization of the chemical composition of soil humic acids using Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta* 153:169–182.

IUSS Working Group WRB, 2014. World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.

Jambu P., Coulibaly G., Bilong P., Magnoux P., Amblès A., 1983. Influence of lipids on the physical properties of soils. In: Novak B (ed) Studies about Humus. Transactions of the VIIth International Symposium. Institute of Crop Protection, Prague, 1:46–50.

Jambu P., Moucawi J., Fustec E., Amblès A., Jacquesy R., 1985. Interrelation entre le pH et la nature des composés lipidiques du sol: étude comparée d'une rendzine et d'un sol lessivé glossique. *Agrochimica* 29:186–198.

Janik L.J., Skjemstad J.O., Shepherd K.D., Spouncer L.R., 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Australian Journal of Soil Research* 45:73–81.

Jansen B., Wiesenberg G.L.B., 2017. Opportunities and limitations related to the application of plant-derived lipid molecular proxies in soil science. *Soil* 3:211–234.

Janzen H.H., 2006. The soil carbon dilemma: Shall we hoard it or use it?. *Soil Biology and Biochemistry* 38:419–424.

Jastrow J.D., Amonette J.E., Bailey V.L., 2007. Mechanisms controlling soil carbon turnover and their potential application for enhancing carbon sequestration. *Climate Change* 80:5–23

Jenny H., 1994. Factors of Soil Formation. A System of Quantitative Pedology. Dover Publications, inc. New York.

Jiménez-Ballesta R., Almendros G., Polo A., Martin de Vidales J.L., 1983. Estudio de un perfil con caracteres isohumicos en la comarca de La Bureba (Burgos). *Anales de Edafología y Agrobiología* 42:1820–1835.

Jiménez-González M.A., De la Rosa J.M., Jiménez-Morillo N.T., Almendros G., González-Pérez J.A., Knicker H., 2016. Post-fire recovery of soil organic matter in a Cambisol from typical Mediterranean forest in Southwestern Spain. *Science of the Total Environment* 572:1414–1421.

Jiménez-González M.A., Álvarez A.M., Carral P., González-Vila F.J., Almendros G., 2017. The diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of soil carbon storage. *Journal of Chromatography A* 1508:130–137.

Jiménez-González M.A., Álvarez A.M., Hernández Z., Almendros G., 2018. Soil carbon storage predicted from the diversity of pyrolytic alkanes. *Biology and Fertility of Soils* 54:617–629.

Jiménez-González M.A., Álvarez A.M., Carral P., Almendros G., 2019. Chemometric assessment of soil organic matter storage and quality from humic acid infrared spectra. *Science of the Total Environment* 685:1160–1168.

Jiménez-Morillo N.T., González-Pérez J.A., Jordán A., Zavala L.M., De la Rosa J.M., Jiménez-González M.A., González-Vila F.J., 2016. Organic matter fractions controlling soil water repellency in sandy soils from the Doñana National Park (Southwestern Spain). *Land Degradation and Development* 27:1413–1423.

Jiménez-Morillo N.T., González-Pérez J.A., Almendros G., De la Rosa J.M., Waggoner D.C., Jordán A., Zavala L.M., González-Vila F.J., Hatcher P.G., 2018. Ultra-high resolution mass spectrometry of physical speciation patterns of organic matter in fire-affected soils. *Journal of Environmental Management* 225:139–147.

Johnson K., Harden J., McGuire A.D., Bliss N.B., Bockheim J.G., Clark M., Nettleton-Hollingsworth T., Jorgenson M.T., Kane E.S., Mack M., O'Donnell J., Ping C-L., Schuur E.A.G., Turetsky M.R., Valentine D.W., 2011. Soil carbon

distribution in Alaska in relation to soil-forming factors. *Geoderma* 167–168:71–84.

Jordán A., Zavala L.M., Mataix-Solera J., Doerr S.H., 2013. Soil water repellency: origin, assessment and geomorphological consequences. *Catena* 108:1–5.

Jordán A., Gordillo-Rivero A.J., García-Moreno J., Zavala L.M., Granged A.J.P., Gil J., Neto-Paixão H.M., 2014. Post-fire evolution of water repellency and aggregate stability in Mediterranean calcareous soils: A 6-year study. *Catena* 118:115–123.

Juo A.S.R., Ayanlaja S.A., Ogunwale J.A., 1976. An evaluation of cation exchange capacity measurements for soils in the tropics. *Communications in Soil Science and Plant Analysis* 7:751–761.

Kamjunke N., von Tümpling W., Hertkorn N., Harir M., Schmitt-Kopplin P., Norf H., Weitere M., Herzsprung P., 2017. A new approach for evaluating transformations of dissolved organic matter (DOM) via high-resolution mass spectrometry and relating it to bacterial activity. *Water Research* 123:513–523.

Nezhad M.T.K., 2019. Storage and drivers of soil organic carbon and nitrogen in a rangeland ecosystem across a lithosequence in western Iran. *Catena* 176:245–263.

Kim S., Kramer R.W., Hatcher P.G., 2003. Graphical method for analysis of ultra-high resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Analytical Chemistry* 75:5336–5344.

Klute A., 1986. Water retention: laboratory methods. In: Klute (Ed.). Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods. American Society of Agronomy and Soil Science Society of America, Madison, pp. 635–660.

Knicker H., 2011. Solid state CPMAS ^{13}C and ^{15}N NMR spectroscopy in organic geochemistry and how spin dynamics can either aggravate or improve spectra interpretation. *Organic Geochemistry* 42:867–890.

Koch B.P., Dittmar T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry* 20:926–932.

Kögel-Knabner I., Zech W., Hatcher P.G., 1988. Chemical composition of the organic matter in forest soils: the humus layer. *Journal of Plant Nutrition and Soil Science* 151:331–340.

Kögel-Knabner I., Hatcher P.G., 1989. Characterization of alkyl carbon in forest soils by CPMAS ^{13}C NMR spectroscopy and dipolar dephasing. *Science of the Total Environment* 81–82:169–177.

Kögel-Knabner I., Guggenberger G., Kleber M., Kandeler E., Kalbitz K., Scheu S., Eusterhues K., Leinweber P., 2008. Organo-mineral associations in temperate soils: Integrating biology, mineralogy, and organic matter chemistry. *Journal of Plant Nutrition and Soil Science* 171:61–82.

Kononova M.M., 1966. Soil Organic Matter: its Nature, its Role in Soil Formation and in Soil Fertility. Pergamon, Oxford

Kononova M.M., 1982. Materia Orgánica del Suelo: su Naturaleza, Propiedades y Métodos de Investigación. Barcelona, Oikos Tau.

Kottek M., Griese J., Beck C., Rudolf B., Rubel F., 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15:259–263.

Kramer R.W., Kujawinski E.B., Hatcher P.G., 2004. Identification of black carbon derived structures in a volcanic ash soil humic acid by Fourier transform ion cyclotron resonance mass spectrometry. *Environmental Science & Technology* 38:3387–3395.

Kruskal J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.

Kumada K., Hurst H.M., 1967. Green humic acid and its possible origin as a fungal metabolite. *Nature* 214:631–633.

Lal R., 2004. Soil carbon sequestration to mitigate climate change. *Geoderma* 123:1–22.

Leinweber P., Schulten H.R., 1999. Advances in analytical pyrolysis of organic matter. *Journal of Analytical and Applied Pyrolysis* 49:359–383.

Legendre P., Legendre L., 1998. Numerical Ecology, vol 24, 2nd edn. Elsevier Sciences, Amsterdam.

Libohova Z., Seybold C., Wysocki D., Wills S., Schoeneberger P., Williams C., Lindbo D., Stott D., Owens P.R., 2018. Reevaluating the effects of soil organic matter and other properties on available water-holding capacity using the National Cooperative Soil Survey Characterization Database. *Journal of Soil and Water Conservation* 73:411–421.

Lobo M.C., Álvarez C., Dorado E., Almendros G., 1984. Recuperacion de los residuos de la vid. Evaluación de su poder fertilizante. In: Recuperación de Recursos de los Residuos. Ingrabel, Soria, pp 681–692.

Lopez-Capel E., Krull E.S., Bol R., Manning D.A., 2008. Influence of recent vegetation on labile and recalcitrant carbon soil pools in central Queensland, Australia: evidence from thermal analysis-quadrupole mass spectrometry-isotope ratio mass spectrometry. *Rapid Communications in Mass Spectrometry* 22:1751–1758.

Lorentz K., Lal R., Ehlers K., 2019. Soil organic carbon stock as an indicator for monitoring land and soil degradation in relation to United Nations' Sustainable Development Goals. *Land Degradation and Development* 30:824–838.

Luinge H.J., van der Maas J.H., Visser T., 1995. Partial least squares regression as a multivariate tool for the interpretation of infrared spectra. *Chemometrics and Intelligent Laboratory Systems* 28:129–138.

Lützow M.V., Kögel-Knabner I., Ekschmitt K., Matzner E., Guggenberger G., Marschner B., Flessa H., 2006. Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions–a review. *European Journal of Soil Science* 57:426–445.

MacCarthy P., Rice J.A., 1985. Spectroscopic methods (other than NMR) for determining functionality in humic substances. In: Aiken, G.R., McKnight, D.M., Wershaw, R.L., MacCarthy, P. (Eds.) Humic Substances in Soils, Sediments and Water, Wiley, New York, pp. 527–559.

Madari B.E., Reeves III J.B., Machado P.L.O.A., Guimarães C.M., Torres E., McCarthy G.W., 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma* 136:245–259.

Manders W.F., 1987. Solid-state ^{13}C NMR determination of the syringyl/guaiacyl ratio in hardwoods. *Holzforschung* 41:13–18.

Manning D.A.C., Renforth P., Lopez-Capel E., Robertson S., Ghazireh N., 2013. Carbonate precipitation in artificial soils produced from basaltic quarry fines and composts: An opportunity for passive carbon sequestration. *International Journal of Greenhouse Gas Control* 17:309-317.

Martín F., Sáiz-Jiménez C., González-Vila F.J., 1979. Pyrolysis–gas chromatography–mass spectometry of lignins. *Holzforschung* 33:210–212.

Martínez A.T., Barrasa J.M., Almendros G., González A.E., 1990. Fungal transformation of lignocellulosics as revealed by chemical and ultrastructural analyses. In: Coughlan, M.P., Collaço, M.T. (Eds.), Advances on Biological Treatments of Lignocellulosic Materials. Elsevier, London, pp. 129–147.

Marynowski L., Smolarek J., Hautevelle Y., 2015. Perylene degradation during gradual onset of organic matter maturation. *International Journal of Coal Geology* 139:17–25.

McKee G.A., Hatcher P.G., 2015. A new approach for molecular characterisation of sediments with Fourier transform ion cyclotron resonance mass spectrometry: Extraction optimisation. *Organic Geochemistry* 85:22–31.

Miller K.E., Lai C.-T., Friedman E.S., Angenent L.T., Lipson D.A., 2015. Methane suppression by iron and humic acids in soils of the Arctic Coastal Plain. *Soil Biology and Biochemistry* 83:176–183.

Miralles I., Piedra-Buena A., Almendros G., González-Vila F.J., González-Pérez J.A., 2015. Pyrolytic appraisal of the lignin signature in soil humic acids: Assessment of its usefulness as carbon sequestration marker. *Journal of Analytical and Applied Pyrolysis* 113:107–115.

Montiel-Rozas M.M., López-García A., Madejón P., Madejón E., 2017. Native soil organic matter as a decisive factor to determine the arbuscular mycorrhizal fungal community structure in contaminated soils. *Biology and Fertility of Soils* 53:327–338.

Muñoz-Rojas M., Román J.R., Roncero-Ramos B., Erickson T.E., Merritt D.J., Aguila-Carricondo P., Cantón Y., 2018. Cyanobacteria inoculation enhances

carbon sequestration in soil substrates used in dryland restoration. *Science of the Total Environment* 636:1149–1154.

Nath T.N., 2014. Soil texture and total organic matter content and its influences on soil water holding capacity of some selected tea growing soils in Sivasagar district of Assam, India. *International Journal of Chemical Sciences* 12:1419–1429.

Nebbioso A., Vinci G., Drosos M., Spaccini R., Piccolo A., 2015. Unveiling the molecular composition of the unextractable soil organic fraction (humin) by humeomics. *Biology and Fertility of Soils* 51:443–451.

Nelson D.V., Sommers L.E., 1982. Total carbon, organic carbon and organic matter. In: Page A.L., Miller R.H., Keeney D.R. (Eds). Methods of Soil Analysis: Part 2, Chemical and Microbiological Properties, 2nd edn. American Society of Agronomy, Madison, pp 539–579.

Nip M., Tegelaar E.W., de Leeuw J.W., Schenck P.A., 1986. A new nonsaponifiable highly aliphatic and resistant biopolymer in plant cuticles. *Naturwissenschaften* 73:579–585.

Oades J.M., 1988. The retention of organic matter in soils. *Biogeochemistry* 5:35–70.

Pan C., Zhao H., Zhao X., Han H., Wang Y., Li J., 2013. Biophysical properties as determinants for soil organic carbon and total nitrogen in grassland salinization. *PLOS ONE* 8.

Panettieri M., Knicker H., Murillo J.M., Madejón E., Hatcher P.G., 2014. Soil organic matter degradation in an agricultural chronosequence under different tillage regimes evaluated by organic matter pools, enzymatic activities and CPMAS ^{13}C NMR. *Soil Biology and Biochemistry* 78:170–181.

Parton W.J., Schimel D.S., Cole C.V., Ojima D.S., 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal* 51:1173–1179.

Pendall E., King J.Y., 2007. Soil organic matter dynamics in grassland soils under elevated CO_2: insights from long-term incubations and stable isotopes. *Soil Biology and Biochemistry* 39:2628–2639.

Pereira P., Rein G., Martín D., 2016. Past and present post-fire environments. *Science of the Total Environment* 573:1275–1277.

Piccolo A., 2002. The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science. *Advances in Agronomy* 75:57–134.

Piccolo A., Spaccini R., Drosos M., Vinci G., Cozzolino V., 2018. The molecular composition of humus carbon: recalcitrance and reactivity in soils. Garcia C., Nannipieri P., Hernandez T. (Eds.), In: The Future of Soil Carbon. Academic Press pp. 87–124.

Pizzeghello D., Francioso O., Concheri G., Muscolo A., Nardi S., 2017. Land use affects the soil C sequestration in Alpine Environment, NE Italy. *Forests* 8:197.

Prince A.L., 1945. Determination of total nitrogen, ammonia, nitrates, and nitrites in soils. *Soil Science* 59:47–52.

Quénéa K., Derenne S., González-Vila F.J., Mariotti A., Rouzaud J.-N., Largeau C., 2005a. Study of the composition of the macromolecular refractory fraction from an acidic sandy forest soil (Landes de Gascogne, France) using chemical degradation and electron microscopy. *Organic Geochemistry* 36:1151–1162.

Quénéa K., Derenne S., Largeau C., Rumpel C., Mariotti A., 2005b. Spectroscopic and pyrolytic features and abundance of the macromolecular refractory fraction in a sandy acid forest soil (Landes de Gascogne, France). *Organic Geochemistry* 36:349–362.

Rabot E., Wiesmeier M., Schlüter S., Vogela H.-J., 2018. Soil structure as an indicator of soil functions: A review. *Geoderma* 314:122–137.

Ralph J., 1999. Lignin Structure: Recent Developments. Madison, WI: US Dairy Forage Research Center, USDA-Agricultural Research Service.

Requena N., Azcón M., Baca M.T., 1996. Chemical changes in humic substances from compost due to incubation with ligno-cellulolytic microorganisms and effects on lettuce growth. *Applied Microbiology and Biotechnology* 45:857–863.

Rhoades J.D., 1982. Cation exchange capacity. In: Methods of Soil Analysis, Part 2 (Page A.L., Miller R.H., Keeney D.R., Eds.) (2nd ed.). Agronomy Monographs 9, ASA and SSSA, Madison, WI. 149–157.

Rumpel C., Seraphin A., Goebel M.O., Wiesenberg G., González-Vila F.J., Bachmann J., Schwark L., Michaelis W., Mariotti A., Kögel-Knabner I., 2004. Alkyl C and hydrophobicity in B and C horizons of an acid forest soil. *Journal of Plant Nutrition and Soil Science* 167:685–692.

Saiz-Jiménez C., de Leeuw J.W., 1986. Lignin pyrolysis products – their structures and their significance as biomarkers. *Organic Geochemistry* 10:869−876.

Sato O., Kumada K., 1967. The chemical nature of the green fraction of P type humic acid. *Soil Science and Plant Nutrition* 13:121–122.

Schmidt M.W.I., Torn M.S., Abiven S., Dittmar T., Guggenberger G., Janssens I.A., Kleber M., Kögel-Knabner I., Lehmann J., Manning D.A.C., Nannipieri P., Rasse D.P., Weiner S., Trumbore S.E., 2011. Persistence of soil organic matter as an ecosystem property. *Nature* 478:49–56.

Schnitzer M., Khan S.U., 1972. Humic Substances in the Environment. Dekker, New York.

Schnitzer M., Khan S.U., 1978. Soil Organic Matter. Elsevier, Amsterdam.

Shahbaz M., Kuzyakov Y., Heitkamp F., 2017. Decrease of soil organic matter stabilization with increasing inputs: Mechanisms and controls. *Geoderma* 304:76–82.

Simoneit B.R.T., Mazurek M.A., 1982. Organic matter of the troposphere—II. Natural background of biogenic lipid matter in aerosols over the rural western United States. *Atmospheric Environment* 16:2139–2159.

Simonetti G., Francioso O., Dal Ferro N., Nardi S., Berti A., Morari F., 2017. Soil porosity in physically separated fractions and its role in SOC protection. *Journal of Soils and Sediments* 17:70–84.

Sleighter R.L., Hatcher P.G., 2007. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *Journal of Mass Spectrometry* 42:559–574.

Sleighter R.L., Mckee G.A., Liu Z., Hatcher P.G., 2008. Naturally present fatty acids as internal calibrants for Fourier transform mass spectra of dissolved organic matter. *Limnology and Oceanography: Methods* 6:246–253.

Smith P., 2008. Land use change and soil organic carbon dynamics. *Nutrient Cycling in Agroecosystems* 81:169–178.

Solomon D., Lehmann J., Thies J., Schäfer T., Liang B., Kinyangi J., Neves E., Petersen J., Luizão F., Skjemstad J., 2007. Molecular signature and sources of biochemical recalcitrance of organic C in Amazonian Dark Earths. *Geochimica et Cosmochimica Acta* 71:2285–2298.

Song X., Li L., Zheng J., Pan G., Zhang X., Zheng J., Hussain Q., Han X., Yu X., 2012. Sequestration of maize crop straw C in different soils: role of oxyhydrates in chemical binding and stabilization as recalcitrance. *Chemosphere* 87:649–654.

Song X.-Y., Liu S.-T., Liu Q.-H., Zhang W.-J., Hu C.-G., 2014. Carbon sequestration in soil humic substances under long-term fertilization in a wheat-maize system from North China. *Journal of Integrative Agriculture* 13:562–569.

Spaccini R., Piccolo A., Conte P., Haberhauer G., Gerzabek M.H., 2002. Increased soil organic carbon sequestration through hydrophobic protection by humic substances. *Soil Biology and Biochemistry* 34:1839–1851.

Spaccini R., Piccolo A., 2012. Carbon sequestration in soils by hydrophobic protection and in situ catalyzed photo-polymerization of soil organic matter (SOM): chemical and physical–chemical aspects of SOM in field plots. In: Piccolo, A. (Ed.), Carbon Sequestration in Agricultural Soils. Springer, Berlin, Heidelberg.

Stevenson F.J., 1994. Humus Chemistry: Genesis, Composition, Reactions. Wiley, New York.

Stewart J.J., Akiyama T., Chapple C., Ralph J., Mansfield S.D., 2009. The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiology* 150:621–635.

Stránský K., Streibl M., Herout V., 1967. On natural waxes. VI. Distribution of wax hydrocarbons in plants at different evolutionary levels. *Collection of Czechoslovak Chemical Communications* 32:3213–3220.

Stubbins A., Spencer R.G.M., Chen H., Hatcher P.G., Mopper K., Hernes P.J., Mwamba V.L., Mangangu A.M., Wabakanghanzi J.N., Six J., 2010. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnology and Oceanography* 55:1467–1477.

Swift R.S., 2001. Sequestration of carbon by soil. *Soil Science* 166:858–871.

Tadini A.M., Constantino I.C., Nuzzo A., Spaccini R., Piccolo A., Moreira A.B., Bisinoti M.C., 2015. Characterization of typical aquatic humic substances in areas of sugarcane cultivation in Brazil using tetramethylammonium hydroxide thermochemolysis. *Science of the Total Environment* 518–519:201–208.

Tian J., Lou Y., Gao Y., Fang H., Liu S., Xu M., Blagodatskaya E., Kuzyakov Y., 2017. Response of soil organic matter fractions and composition of microbial community to long-term organic and mineral fertilization. *Biology and Fertility of Soils* 53:523–532.

Tinoco P., Almendros G., González-Vila F.J., 2002. Impact of the vegetation on the lignin pyrolytic signature of soil humic acids from Mediterranean soils. *Journal of Analytical and Applied Pyrolysis* 64:407–420.

Tinoco P., Almendros G., González-Vila F.J., Sanz J., González-Pérez J.A., 2015. Revisiting molecular characteristics responsive for the aromaticity of soil humic acids. *Journal of Soils and Sediments* 15:781–791.

Tinoco P., Almendros G., Sanz J., 2018. Soil perturbation in Mediterranean ecosystems reflected by differences in free-lipid biomarker assemblages. *Journal of Agricultural and Food Chemistry* 66:9895–9906.

Towett E.K., Shepherd K.D., Tondoh J.E., Winowiecki L.A., Lulseged T., Nyambura M., Sila A., Vågen T.-G., Cadisch G., 2015. Total elemental composition of soils in Sub-Saharan Africa and relationship with soil forming factors. *Geoderma Regional* 5:157–168.

Traina S.J., Novak J., Smeck N.E., 1990. An ultraviolet absorbance method of estimating the percent aromatic carbon content of humic acids. *Journal of Environmental Quality* 19: 151–153.

Van Krevelen D.W., 1950. Graphical-statistical method for the study of structure and reaction processes of coal. *Fuel* 29:269–284.

Vanholme R., Demedts B., Morreel K., Ralph J., Boerjan W., 2010. Lignin biosynthesis and structure. *Plant Physiology* 153:895–905.

Viscarra Rossel R.P., 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems* 90:72–83.

Waggoner D.C., Chen H., Willoughby A.S., Hatcher P.G., 2015. Formation of black carbon-like and alicyclic aliphatic compounds by hydroxyl radical initiated degradation of lignin. *Organic Geochemistry* 82:69–76.

Walkley A., Black I.A., 1934. An examination of degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science* 37:29–37.

Wenting F., Plante A.F., Aufdenkampe A.K., Six J., 2014. Soil organic matter stability in organo-mineral complexes as a function of increasing C loading. *Soil Biology and Biochemistry* 69:398–405.

Wold S., Sjöström M., Eriksson L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58:109–130.

Wu G.-L., Liu Z.-H., Zhang L., Hu T.-M., Chen J.-M., 2010. Effects of artificial grassland establishment on soil nutrients and carbon properties in a black-soil-type degraded grassland. *Plant and Soil* 333:469–479.

Wu Z., Dijkstra P., Koch G.W., Peñuelas J., Hungate B.A., 2011. Responses of terrestrial ecosystems to temperature and precipitation change: a meta-analysis of experimental manipulation. *Global Change Biology* 17:927–942.

Yang Q., Wu S., Lou R., LV G., 2011. Structural characterization of lignin from wheat straw. *Wood Science and Technology* 45:419–431.

Yazdanshenas H., Tavili A., Mohammad J., Shafeian E., 2018. Evidence for relationship between carbon storage and surface cover characteristics of soil in rangelands. *Catena* 167:139–146.

Yuan H., Ge T., Chen C., O'Donnell A.G., Wu J., 2012. Significant role for microbial autotrophy in the sequestration of soil carbon. *Applied and Environmental Microbiology* 78:2328–2336.

Zech W., Haumaier L., Kögel-Knabner I., 1989. Changes in aromaticity and carbon distribution of soil organic matter due to pedogenesis. *Science of the Total Environment* 81–82:179–186.

Anexo I:

Tablas de Resultados y Figuras

Table 1. Sampling points and soil classification

Sample No.	Geographic coordinates	Elevation (m a.s.l.)	Soil type (FAO, 2014)	Soil type (USDA, 2014)
1	40°33'N 4°8'W	1150	Dystric Cambisol (Humic)	Humixerept
2	41°7'N 3°34'W	1580	Haplic Umbrisol (Hyperhumic)	Humixerept
3	40°23'N 3°16'W	870	Calcaric Cambisol (Humic)	Humixerept
4	40°53'N 3°34'W	950	Gleyic Cambisol (Humic)	Humaquept
5	40°53'N 3°34'W	950	Dystric Cambisol (Humic)	Humixerept
6	42°36'N 3°11'W	830	Calcic Chernozem (Pachic)	Calcixeroll
7	40°44'N 3°42'W	530	Dystric Cambisol (Ochric)	Dystroxerept
8	40°47'N 2°57'W	1060	Leptic Kastanozem (Hyperhumic)	Calcixeroll
9	41°14'N 3°24'W	1600	Leptic Podzol (Arenic)	Haplorthod
10	40°44'N 3°48'W	1000	Dystric Cambisol (Humic)	Haploxerept
11	40°21'N 3°56'W	600	Dystric Cambisol (Humic)	Haploxerept
12	40°33'N 3°43'W	700	Dystric Cambisol (Loamic)	Haploxerept
13	40°54'N 3°28'W	1000	Eutric Cambisol (Humic)	Haploxerept
14	40°52'N 3°34'W	1000	Eutric Cambisol (Humic)	Haploxerept
15	40°58'N 3°37'W	1000	Dystric Cambisol (Humic)	Haploxerept
16	40°54'N 3°53'W	1187	Dystric Cambisol (Humic)	Dystroxerept
17	40°51'N 3°44'W	1316	Dystric Cambisol (Colluvic)	Dystroxerept
18	40°45'N 3°41'W	940	Leptic Cambisol (Humic)	Haploxerept
19	43°15'N 2°51'W	375	Leptic Umbrisol (Loamic)	Hapludoll
20	43°4'N 2°35'W	830	Haplic Luvisol (Humic)	Hapludalf
21	42°34'N 2°38'W	580	Eutric Cambisol (Humic)	Haploxerept
22	43°15'N 2°51'W	350	Haplic Umbrisol (Loamic)	Dystrudept
23	42°28'N 8°53'W	200	Leptic Regosol (Humic)	Udorthent
24	42°36'N 8°38'W	150	Leptic Regosol (Humic)	Udorthent
25	43°4'N 8°22'W	350	Leptic Umbrisol (Hyperhumic)	Humudept
26	28°22'N 16°39'W	364	Vitric Andosol (Hyperhumic)	Ustivitrand
27	28°26'N 16°29'W	150	Leptic Regosol (Arenic)	Xerorthent
28	28°14'N 16°28'W	882	Leptic Regosol (Arenic)	Udorthent
29	28° 9'N 16°38'W	1469	Folic Umbrisol (Chromic)	Haplustept
30	40°13'N 4°29'W	765	Dystric Regosol (Arenic)	Xerorthent
31	41°29'N 4°19'W	835	Eutric Cambisol (Humic)	Haploxerept
32	40°18'N 4°38'W	670	Eutric Cambisol (Arenic)	Haploxerept
33	41°1'N 3°12'W	1010	Eutric Cambisol (Humic)	Haploxerept
34	40°58'N 3°44'W	1195	Eutric Cambisol (Humic)	Haploxerept
35	40°56'N 3°41'W	1200	Dystric Leptosol (Humic)	Xerorthent

Table 2. General characteristics of sampling sites

Sample No.	Vegetation	Vegetation class	Geological substrate	Climate type (Köppen)	Emberger's *Q* index
1	*Quercus pyrenaica*	Angiosperm	Granite	Csa	89
2	*Pinus sylvestris*	Gymnosperm	Gneiss	Csa	160
3	*Quercus ilex*	Angiosperm	Limestone	Csa	40
4	*Fraxinus angustifolia*	Angiosperm	Granite	Csa	78
5	*Paeonia coriacea,*	Grass	Granite	Csa	78
6	Pastureland with: *Micropyrum tenellum, Trifolium dubium, Trifolium campestre, Anacyclus clavatus*	Grass	Limestone	Csa	65
7	*Quercus rotundifolia*	Angiosperm	Gneiss	Csa	76
8	*Quercus rotundifolia*	Angiosperm	Limestone	Csa	62
9	*Fagus sylvatica*	Angiosperm	Schist	Csa	160
10	*Pinus sylvestris*	Gymnosperm	Granite	Csa	76
11	*Pinus pinea*	Gymnosperm	Sandstone	Csa	38
12	*Quercus rotundifolia*	Angiosperm	Sandstone	Csa	47
13	*Juniperus oxycedrus*	Gymnosperm	Schist	Csa	89
14	*Juniperus oxycedrus*	Gymnosperm	Gneiss	Csa	89
15	*Pinus pinaster*	Gymnosperm	Gneiss	Csa	89
16	*Quercus pyrenaica*	Angiosperm	Gneiss	Csa	89
17	*Pinus sylvestris*	Gymnosperm	Granite	Csa	89
18	*Quercus ilex*	Angiosperm	Granite	Csa	78
19	Pastureland for grazing: *Brachypodium retusum, Lolium perenne, Trifolium repens*	Grass	Sandstone	Cfb	278
20	*Fagus sylvatica*	Angiosperm	Limestone	Cfb	143
21	Pastureland for grazing: *Brachypodium retusum, Cynosurus cristatus, Trifolium repens*	Grass	Limestone	Cfb	126
22	*Pinus radiata*	Gymnosperm	Sandstone	Cfb	126
23	*Pinus pinaster*	Gymnosperm	Granite	Cfa	278
24	*Pinus pinaster*	Gymnosperm	Granite	Cfa	187
25	*Pinus pinaster*	Gymnosperm	Schist	Cfa	187
26	*Laurus canariensis*	Angiosperm	Volcanic	Csa	102
27	*Euphorbia canariensis*	Angiosperm	Volcanic	Csa	84
28	Fallow: *Solanum tuberosum*	Grass	Volcanic	Csa	84
29	*Pinus canariensis*	Gymnosperm	Volcanic	Csa	71
30	*Pinus pinea*	Gymnosperm	Sandstone	Csa	52
31	*Pinus pinea*	Gymnosperm	Limestone	Csa	52
32	*Quercus rotundifolia*	Angiosperm	Granite	Csa	78
33	*Quercus rotundifolia*	Angiosperm	Schist	Csa	62
34	*Juniperus thurifera*	Gymnosperm	Gneiss	Csa	89
35	*Juniperus thurifera*	Gymnosperm	Limestone	Csa	89

Table 3. Soil physical properties

Sample No.	Sand ($g \cdot kg^{-1}$)	Silt ($g \cdot kg^{-1}$)	Clay ($g \cdot kg^{-1}$)	Texture	Bulk density ($g \cdot cm^{-3}$)	WHC ($g \cdot kg^{-1}$)	Wilting point ($g \cdot kg^{-1}$)
1	581	261	158	Sandy loam	0.96	619	182
2	603	320	77	Sandy loam	0.89	651	187
3	215	758	27	Silt loam	0.81	813	325
4	622	264	113	Sandy loam	0.85	893	296
5	522	352	126	Sandy loam	0.86	657	174
6	502	240	258	Sandy clay loam	1.21	506	201
7	768	156	76	Sandy loam	1.13	427	72
8	267	415	318	Clay loam	0.77	832	333
9	593	318	89	Sandy loam	1.06	510	102
10	768	116	116	Sandy loam	0.62	1122	258
11	828	56	116	Loamy sand	0.76	842	376
12	848	56	96	Loamy sand	0.86	850	242
13	548	276	176	Sandy loam	0.56	1272	466
14	648	216	136	Sandy loam	0.67	1007	476
15	788	116	96	Sandy loam	0.75	910	271
16	620	180	200	Sandy clay loam	0.89	676	174
17	540	290	170	Sandy loam	1.09	479	125
18	380	410	210	Loam	0.68	1206	497
19	384	374	242	Loam	0.90	623	190
20	169	537	294	Silty clay loam	0.92	704	218
21	326	291	383	Clay loam	0.82	942	367
22	386	403	211	Loam	0.95	612	166
23	630	210	160	Sandy loam	0.65	1124	270
24	670	220	110	Sandy loam	0.79	782	190
25	320	420	260	Loam	0.54	1420	383
26	610	290	100	Sandy loam	0.98	539	169
27	420	300	280	Clay loam	1.06	600	211
28	600	230	170	Sandy loam	0.99	599	202
29	680	140	180	Sandy loam	1.01	894	276
30	935	27	38	Sand	1.11	467	131
31	913	30	58	Sand	0.64	991	396
32	717	149	135	Sandy loam	0.94	567	106
33	378	488	134	Silt loam	0.64	970	376
34	637	167	196	Sandy loam	0.59	1196	742
35	392	384	224	Loam	0.91	742	347

WHC, water holding capacity (atmospheric pressure).

Table 4. Soil chemical properties

Sample No.	SOC ($g \cdot kg^{-1}$)	C/N	Exchangeable cations ($cmol_c \cdot kg^{-1}$)					pH	EC ($dS \cdot m^{-1}$)
			Na^+	K^+	Ca^{2+}	Mg^{2+}	CEC		
1	41	11.3	0.26	0.37	6.14	2.83	16.4	5.2	0.488
2	67	14.8	0.26	0.31	0.85	0.31	24.4	3.91	0.385
3	96	15.3	0.38	3.02	26.87	4.96	30.8	7.16	1.143
4	87	13.3	0.29	0.90	14.36	3.13	22.0	6.24	0.747
5	48	13.1	0.36	0.31	6.16	1.28	14.4	5.27	0.418
6	17	13.9	0.36	0.93	25.06	2.11	15.9	7.63	1.025
7	18	16.0	0.23	0.24	1.88	0.50	4.5	5.2	0.290
8	87	13.3	0.16	1.23	30.95	1.73	21.5	6.82	1.326
9	32	16.4	0.27	0.48	2.85	0.62	11.6	5.71	0.490
10	140	18.1	0.17	0.96	8.99	2.82	14.2	5.13	0.549
11	117	26.7	0.20	0.66	7.69	3.78	13.9	4.9	0.499
12	93	8.9	0.15	0.92	16.38	2.41	20.1	6.4	0.693
13	134	12.1	0.20	2.47	29.16	7.39	19.8	6.99	1.356
14	104	18.5	0.17	0.92	19.85	4.22	22.2	6.45	0.786
15	81	16.9	0.17	0.73	10.20	3.61	17.1	5.99	0.552
16	55	18.0	0.24	0.73	6.35	2.53	15.9	5.72	0.309
17	39	13.0	0.30	0.55	4.28	1.28	13.8	5.57	0.186
18	105	17.0	0.44	2.54	39.79	4.09	41.9	7.18	0.767
19	41	15.8	0.32	0.26	3.20	0.38	15.8	4.64	0.381
20	44	13.4	0.13	0.33	9.30	1.17	11.3	5.14	0.491
21	57	13.9	0.30	0.92	23.65	5.68	32.8	6.88	0.645
22	27	17.0	0.20	0.10	1.13	0.21	13.3	4.18	0.324
23	133	31.0	0.31	0.35	0.34	0.54	32.4	3.49	0.485
24	90	20.0	0.28	0.24	0.45	0.61	22.6	3.73	0.361
25	132	18.0	0.34	0.27	0.86	0.89	31.7	4.23	0.527
26	18	22.0	1.56	1.41	10.47	7.04	28.3	6.76	0.514
27	22	12.0	0.89	3.44	13.21	3.11	18.3	7.69	0.850
28	23	14.0	2.32	6.59	10.64	7.96	29.1	7.29	0.784
29	105	27.0	0.63	1.54	9.06	4.13	20.2	6.41	0.475
30	35	20.0	0.15	0.34	3.08	1.04	4.9	5.74	0.327
31	99	25.9	0.17	0.62	8.41	2.70	9.3	5.74	0.561
32	46	16.8	0.29	0.52	5.31	1.47	12.8	5.4	0.376
33	89	23.7	0.21	1.17	12.44	3.47	17.2	6.2	0.759
34	157	21.6	0.16	1.74	17.33	4.32	18.0	6.55	1.087
35	92	13.9	0.19	0.69	23.20	4.49	25.9	7.4	0.820

SOC, soil organic carbon; *CEC*, cation exchange capacity*; EC*, electrical conductivity.

Table 5. Soil organic matter fractions

	Organic fractions (g C· 100 C g $^{-1}$ soil)				
Sample No.	**FOM**	**FA**	**HA**	**Humin**	**HA/FA**
1	3.5	9.9	46.0	35.4	4.64
2	6.6	10.8	47.1	28.3	4.35
3	3.4	11.0	31.5	49.7	2.87
4	6.9	15.6	35.4	38.2	2.27
5	8.1	9.3	38.3	40.5	4.14
6	10.7	4.1	34.9	45.0	8.60
7	3.9	12.3	28.0	49.8	2.27
8	5.4	5.5	37.0	46.4	6.68
9	5.2	16.0	28.2	42.9	1.76
10	13.8	8.0	26.5	49.1	3.30
11	23.7	5.0	13.0	56.7	2.58
12	5.1	12.6	35.7	44.1	2.84
13	14.0	10.0	39.3	31.8	3.95
14	16.3	8.1	36.3	36.6	4.46
15	14.6	13.5	38.9	30.0	2.89
16	10.8	11.9	30.3	44.0	2.54
17	8.8	14.0	47.0	24.8	3.35
18	8.5	9.1	36.6	42.8	4.00
19	4.7	10.7	46.8	29.9	4.37
20	9.1	15.1	34.0	34.2	2.25
21	7.0	17.5	48.7	21.1	2.78
22	6.0	12.4	38.0	32.1	3.07
23	15.0	12.0	43.3	25.3	3.61
24	7.5	11.3	48.8	26.9	4.30
25	8.9	11.6	39.1	31.9	3.37
26	0.7	5.4	51.8	34.7	9.57
27	12.6	8.0	31.5	42.6	3.92
28	9.3	7.4	38.4	40.8	5.15
29	12.3	9.2	27.1	48.5	2.95
30	17.4	12.6	21.3	45.5	1.68
31	23.4	7.1	10.1	56.6	1.43
32	5.1	9.3	44.7	37.2	4.81
33	23.5	11.8	18.0	41.6	1.52
34	24.9	6.3	12.1	54.4	1.94
35	3.4	13.9	37.5	40.3	2.70

FOM, free organic matter*; FA*, fulvic acid; *HA*, humic acid.

Table 6. Spectroscopic parameters in the visible range and elementary composition of the humic acids

Sample No.	E4 (AU)	E6 (AU)	E4/E6	DHPQ (AU·10⁴)	Elementary composition (g·100 g⁻¹)				
					C	H	N	S	O
1	0.78	0.15	5.19	2.48	54.3	5.0	5.7	0.7	34.3
2	1.14	0.20	5.82	1.38	55.9	4.4	4.6	0.3	34.8
3	0.66	0.09	7.16	0.82	57.2	4.8	4.1	0.6	33.3
4	0.38	0.06	6.70	0.36	56.8	5.8	5.5	0.7	31.2
5	0.51	0.08	6.15	0.60	56.4	5.5	5.7	0.9	31.4
6	1.29	0.25	5.22	0.91	53.7	3.8	4.6	0.7	37.2
7	0.68	0.11	6.18	1.74	56.0	5.6	4.3	0.4	33.7
8	0.86	0.16	5.22	0.59	57.3	4.7	4.3	0.4	33.2
9	0.78	0.13	5.86	3.93	57.8	5.2	4.9	0.4	31.7
10	0.71	0.11	6.25	1.10	57.6	5.1	3.2	0.3	33.8
11	0.59	0.07	8.82	0.60	57.2	4.8	3.4	0.4	34.3
12	0.53	0.08	6.71	0.47	55.8	5.0	4.8	0.4	33.9
13	0.51	0.06	8.64	0.62	56.3	5.1	3.1	0.5	35.0
14	0.61	0.07	8.35	0.44	52.3	4.6	2.9	0.4	39.8
15	0.48	0.07	6.93	0.29	59.1	5.6	3.6	0.4	31.3
16	0.55	0.09	6.35	0.94	57.5	5.3	3.8	0.4	33.0
17	0.80	0.15	5.21	1.65	54.9	4.8	4.9	0.6	34.8
18	0.51	0.07	7.36	0.53	55.0	4.8	4.4	0.5	35.2
19	0.95	0.20	4.87	2.61	54.1	4.5	3.7	0.4	37.3
20	0.61	0.10	6.27	2.12	55.8	5.1	4.4	0.5	34.2
21	0.66	0.10	6.44	0.66	55.0	4.6	3.7	0.4	36.3
22	0.95	0.22	4.33	2.71	52.9	4.2	3.4	0.3	39.2
23	1.18	0.25	4.66	1.32	57.8	4.0	4.2	0.3	33.8
24	0.76	0.14	5.52	0.37	56.9	4.6	3.2	0.3	35.0
25	0.76	0.14	5.58	0.78	54.3	4.8	4.2	0.4	36.3
26	1.60	0.36	4.48	2.65	50.9	3.3	4.3	0.5	41.0
27	0.67	0.11	6.13	1.09	57.8	4.9	5.4	0.7	31.2
28	0.72	0.14	5.10	0.35	50.5	4.3	5.0	0.8	39.4
29	0.55	0.09	6.08	0.34	51.6	4.4	3.0	0.3	40.8
30	0.66	0.12	5.55	0.66	58.3	5.5	4.1	0.4	31.8
31	0.58	0.08	6.85	0.78	58.3	5.2	3.1	0.4	33.0
32	1.13	0.23	4.81	1.00	56.0	4.3	3.6	0.3	35.7
33	0.64	0.10	6.32	0.64	57.9	5.1	3.1	0.3	33.5
34	0.46	0.06	8.28	0.46	57.0	5.0	3.7	0.4	33.9
35	0.62	0.09	6.52	0.67	55.0	4.8	4.3	0.4	35.4

E4, optical density of HA at 465 nm (0.2 mg·cm^{-3}); *E6*, optical density of HA at 665 nm (0.2 mg·cm^{-3}); DHPQ, peak intensity at 616 and 572 nm measured in the 2nd derivative spectrum.

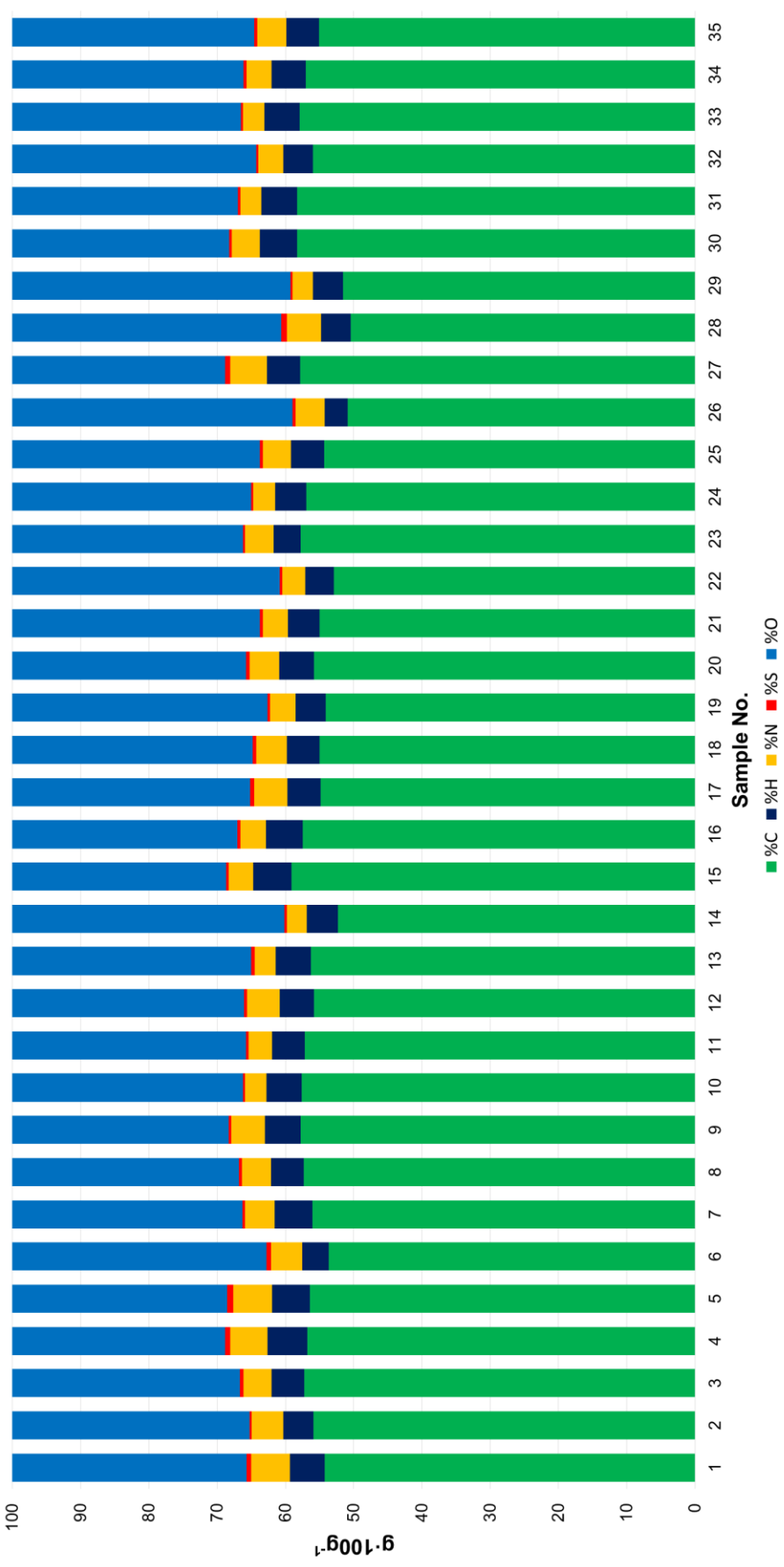


Figure 1. Results of the elemental analysis of humic acids indicated in Table 6.

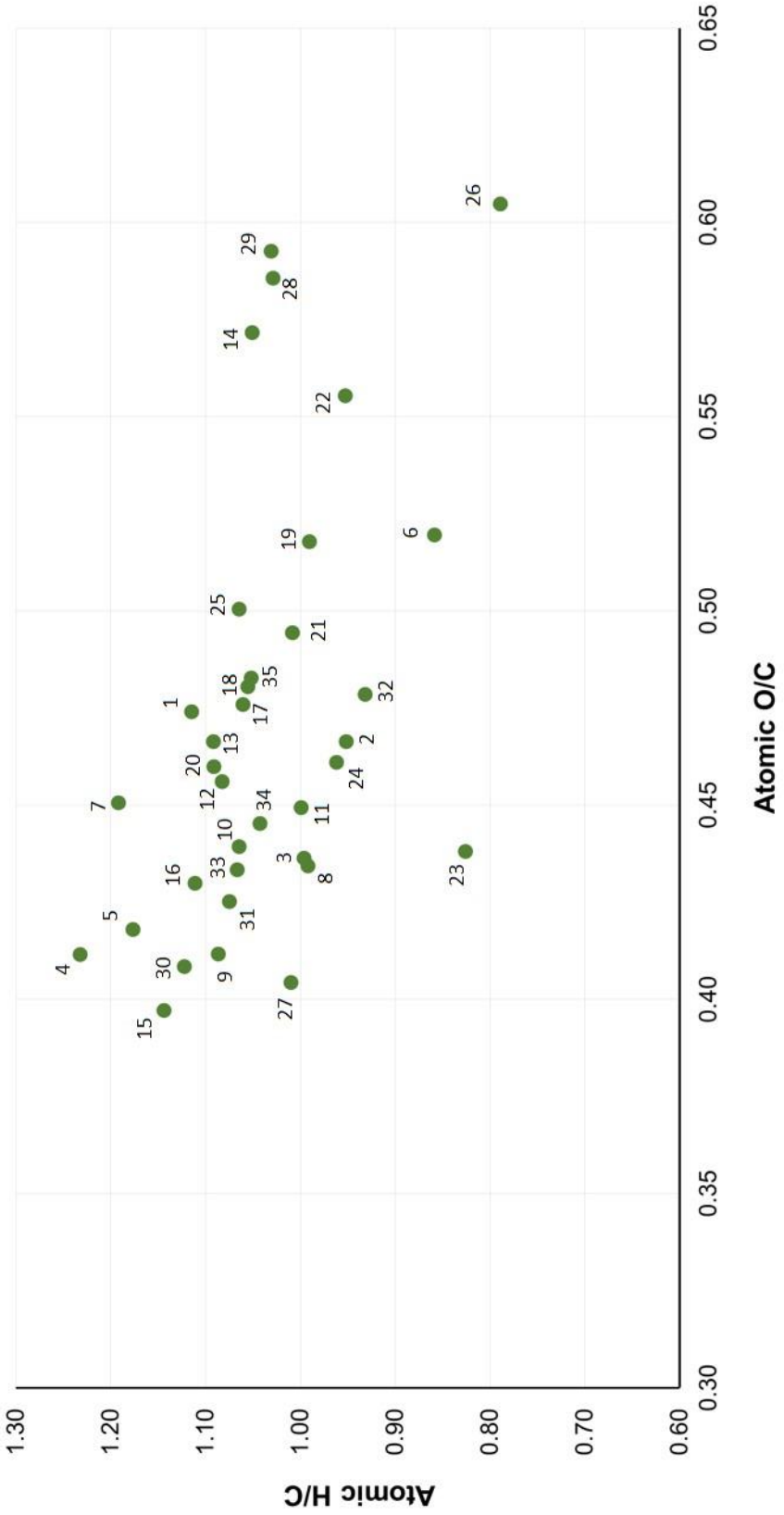


Figure 2. Van Krevelem diagram of humic acids from the soils.

Table 7. Peak area integration values of the ^{13}C NMR spectra of humic acids

Sample No.	Alkyl C (%) (0–45 ppm)	*N*-alkyl + OCH_3 C (%) (45–60 ppm)	*O*-alkyl C (%) (60–110 ppm)	Aromatic C (%) (110–160 ppm)	Carbonyl C (%) (160–220 ppm)	Het-/H-Ar
1	28.4	10.8	27.4	18.9	14.6	0.33
2	32.7	8.4	18.7	27.8	12.5	0.30
3	31.4	11.3	24.2	21.9	11.3	0.40
4	35.4	11.7	24.8	16.8	11.3	0.37
5	31.6	11.8	24.8	19.1	12.7	0.34
6	30.8	9.6	14.3	32.3	13.0	0.26
7	41.1	9.7	21.0	18.1	10.1	0.34
8	32.9	9.9	22.2	22.1	12.9	0.35
9	37.2	10.4	21.4	18.1	12.9	0.34
10	32.4	9.6	23.1	24.1	10.9	0.39
11	31.0	11.0	25.2	22.4	10.5	0.42
12	33.2	10.7	26.6	16.6	12.8	0.40
13	36.0	9.9	24.6	18.1	11.5	0.41
14	31.0	9.4	26.3	21.0	12.2	0.45
15	34.8	10.2	25.2	19.2	10.6	0.40
16	36.1	10.2	25.7	17.4	10.7	0.40
17	33.9	10.5	25.3	18.9	11.5	0.30
18	32.3	11.8	24.7	18.6	12.7	0.40
19	37.0	8.9	20.1	22.6	11.4	0.30
20	33.8	11.2	25.7	18.2	11.1	0.36
21	31.3	10.3	26.5	19.1	12.8	0.42
22	41.5	8.8	18.6	20.6	10.5	0.30
23	32.0	8.8	20.2	26.4	12.7	0.33
24	35.0	8.9	22.3	21.7	12.0	0.38
25	29.7	10.3	26.0	21.9	12.1	0.36
26	25.5	7.9	16.8	35.1	14.8	0.27
27	37.0	10.9	19.2	20.4	12.5	0.36
28	31.9	11.3	21.7	22.6	12.5	0.36
29	32.9	10.1	22.5	24.1	10.4	0.35
30	39.5	10.5	21.4	18.4	10.3	0.32
31	36.8	10.2	22.4	20.8	9.8	0.36
32	31.5	8.7	21.3	25.0	13.6	0.35
33	35.4	10.3	24.1	20.3	9.8	0.38
34	37.3	10.9	23.7	18.8	9.4	0.38
35	34.8	11.0	25.0	17.8	11.5	0.35

Het-/H-Ar, ratio Heteroaromatic / H-aromatic C (140–160 ppm/110–140 ppm).

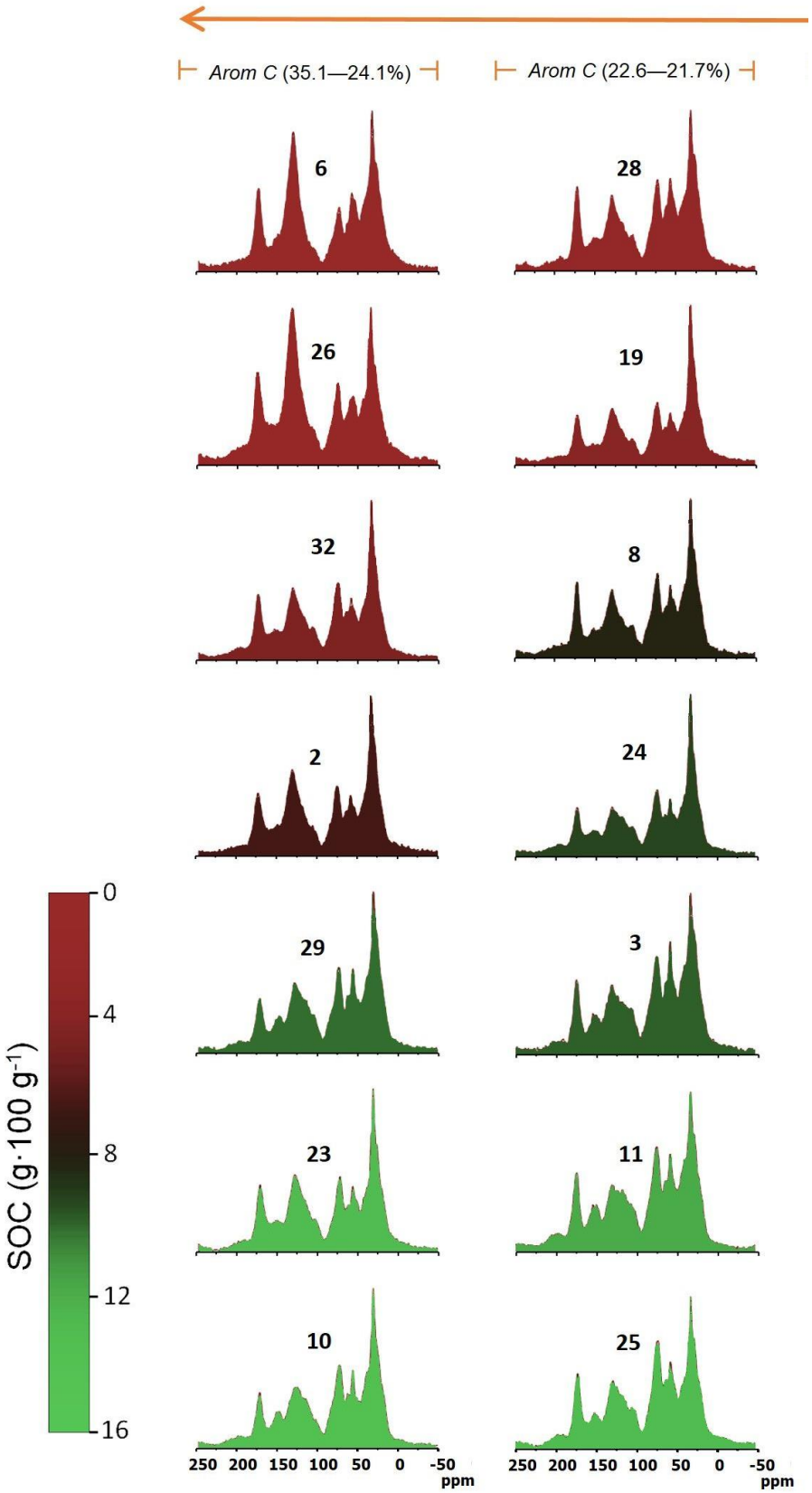


Figure 3. ^{13}C NMR spectra of humic acids from Mediterranean soils. Samples are ordered horizontally in function of the aromatic C content (Table 7) and vertically in function of increasing content of SOC (Table 4). The samples are colored according the SOC content. Numbers on the spectra refer to Table 1.

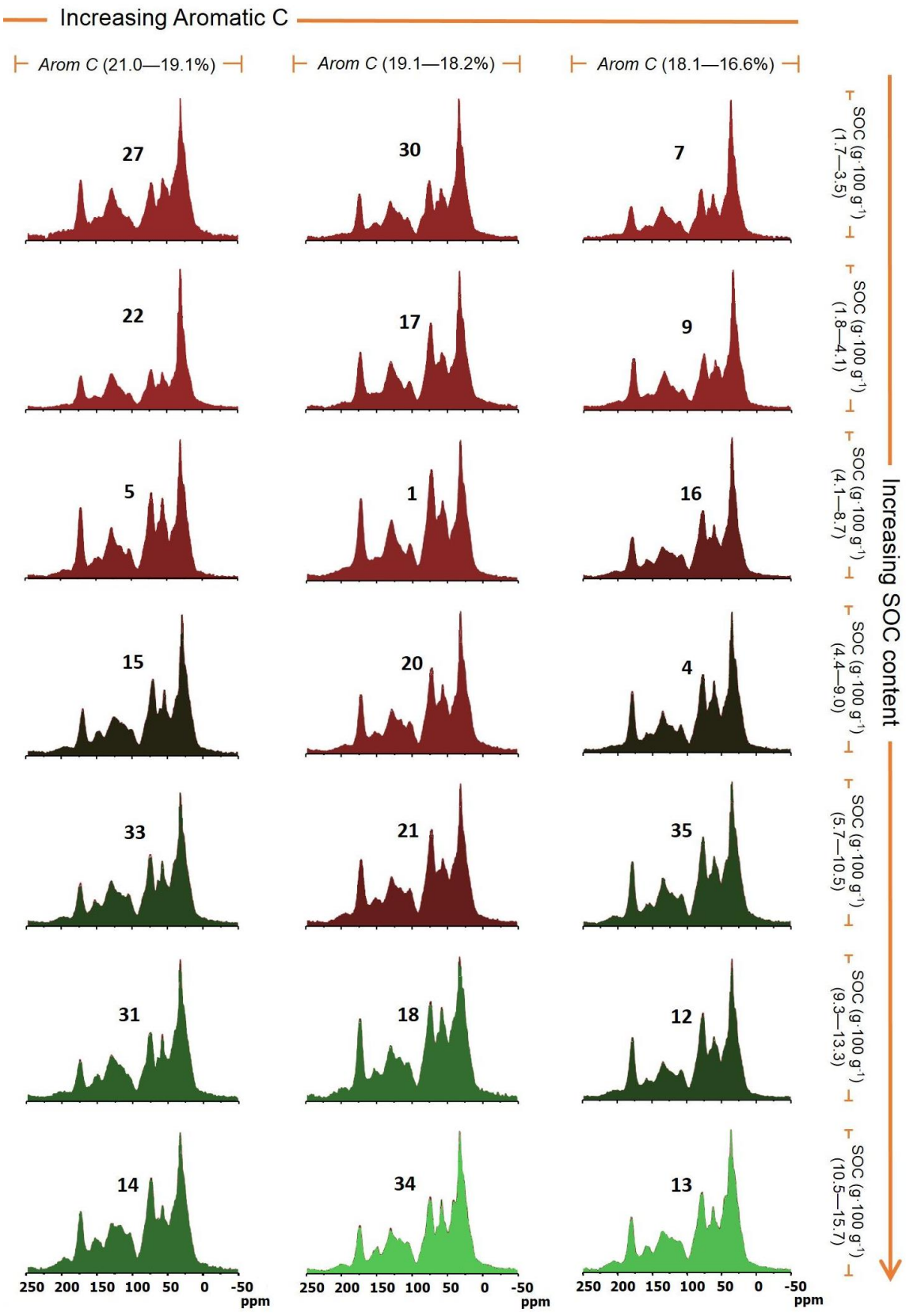

Increasing Aromatic C
Arom C (21.0—19.1%)
Arom C (19.1—18.2%)
Arom C (18.1—16.6%)
27
30
7
22
17
9
5
1
16
15
20
4
33
21
35
31
18
12
14
34
13
250 200 150 100 50 0 -50
ppm
SOC (g·100 g-1) (1.7—3.5)
SOC (g·100 g-1) (1.8—4.1)
SOC (g·100 g-1) (4.1—8.7)
SOC (g·100 g-1) (4.4—9.0)
SOC (g·100 g-1) (5.7—10.5)
SOC (g·100 g-1) (9.3—13.3)
SOC (g·100 g-1) (10.5—15.7)
Increasing SOC content

Table 8. Shannon *H'* diversity index and ratios calculated from the abundances of *n*-alkanes released by analytical pyrolysis from whole soil samples

Sample No.	RLC ($<C_{20}/>C_{20}$)	CPI ($2n+1/2n$)	*H'* total alkanes	*H'* ($<C_{20}$) alkanes	*H'* ($>C_{20}$) alkanes	*H'* ($2n$) alkanes	*H'* ($2n+1$) alkanes
1	0.82	1.16	2.84	2.33	2.01	2.08	2.21
2	1.84	1.37	2.99	2.48	2.09	2.25	2.36
3	1.70	1.22	2.96	2.48	1.98	2.22	2.31
4	0.70	2.03	2.90	2.46	2.05	2.31	2.24
5	1.38	1.08	3.00	2.47	2.12	2.26	2.35
6	3.33	1.01	2.70	2.19	1.91	1.93	2.07
7	0.56	1.66	2.86	2.44	2.08	2.33	2.11
8	2.22	1.33	2.98	2.47	2.11	2.21	2.36
9	1.17	1.00	2.71	2.20	1.82	1.97	2.07
10	4.05	1.30	2.90	2.44	2.21	2.15	2.26
11	2.97	1.41	2.94	2.42	2.26	2.22	2.30
12	1.29	1.50	2.97	2.47	2.04	2.26	2.32
13	1.45	1.66	3.00	2.46	2.13	2.26	2.39
14	2.06	1.55	3.00	2.47	2.15	2.22	2.39
15	1.23	1.55	3.04	2.47	2.21	2.32	2.41
16	0.56	1.90	2.80	2.45	1.97	2.35	2.05
17	0.97	1.67	3.06	2.47	2.27	2.36	2.42
18	1.98	1.33	3.03	2.48	2.21	2.25	2.41
19	3.79	1.03	2.85	2.48	1.91	2.10	2.21
20	3.98	1.05	2.92	2.47	2.20	2.17	2.28
21	2.63	1.05	3.01	2.46	2.31	2.24	2.39
22	0.27	2.56	2.42	2.47	1.76	2.17	1.70
23	1.21	1.55	3.02	2.48	2.16	2.32	2.37
24	1.78	1.28	3.04	2.48	2.23	2.28	2.42
25	2.57	1.27	2.98	2.46	2.22	2.24	2.34
26	0.39	1.51	2.65	2.24	1.91	2.00	1.94
27	6.16	1.08	2.83	2.36	1.87	2.10	2.20
28	2.46	1.09	2.78	2.45	1.50	2.03	2.13
29	3.16	1.23	2.96	2.46	2.22	2.21	2.32
30	0.93	1.68	3.04	2.45	2.25	2.34	2.40
31	1.25	1.04	2.99	2.41	2.16	2.25	2.34
32	1.16	1.42	3.06	2.46	2.26	2.32	2.43
33	0.58	1.57	2.69	2.27	1.90	2.00	2.04
34	0.91	1.90	3.01	2.47	2.19	2.35	2.38
35	1.72	1.30	3.02	2.47	2.19	2.26	2.40

RLC, relative chain length (short-chain to long-chain); *CPI*, carbon preference index (odd-to-even-C numbered); *H'*, Shannon index.

Table 9. Shannon *H'* diversity index and ratios calculated from total abundances of methoxyphenols released by analytical pyrolysis

Sample	*H'* Guaiacyl	*H'* Syringyl	*H'* Total	C_{0-2}-MeO-Ph/ C_3-MeO-Ph
1	1.484	1.622	1.935	12.477
2	1.569	1.646	1.770	13.820
3	1.627	1.688	2.173	6.508
4	1.559	1.683	2.166	4.406
5	1.399	1.607	1.858	10.240
6	1.582	1.456	2.202	5.530
7	1.563	1.690	1.912	18.450
8	1.632	1.724	2.207	4.649
9	1.615	1.690	1.990	6.745
10	1.652	1.713	1.708	6.515
11	1.617	1.754	1.735	5.963
12	1.593	1.664	2.154	5.763
13	1.596	1.699	1.853	7.701
14	1.646	1.744	1.972	5.828
15	1.638	1.731	1.937	6.714
16	1.583	1.598	1.966	7.380
17	1.510	1.557	1.796	13.584
18	1.619	1.718	2.186	7.340
19	1.543	1.709	1.849	13.600
20	1.294	1.500	1.476	23.934
21	1.269	1.505	1.535	25.052
22	1.543	1.377	1.712	17.066
23	1.572	1.718	1.755	12.623
24	1.520	1.692	1.740	15.551
25	1.564	1.675	1.751	12.451
26	1.524	1.660	2.249	4.532
27	1.349	1.163	1.949	30.005
28	1.127	1.118	1.266	34.015
29	1.606	1.702	1.890	8.134
30	1.446	1.714	1.530	17.439
31	1.620	1.739	1.731	5.962
32	1.350	1.701	1.592	40.682
33	1.590	1.678	2.092	5.689
34	1.655	1.722	1.983	4.965
35	1.657	1.729	1.972	4.605

H', Shannon index; C_{0-2}-MeO-Ph/C_3-MeO-Ph, ratio between short-chain to long-chain methoxyphenols (G+MG+EG+VG+AG+S+MS+ES+VS+AS)/(PG+PS).

Table 10. Pyrolysis products identified in soil samples with an indication of the diagnostic value of each of the compounds with respect to the different formation factors studied

Compound	Environmental proxy value[a]
Alkanes	
n-Alkane C_9	− + −
n-Alkane C_{10}	− − −
n-Alkane C_{11}	− − −
n-Alkane C_{12}	− − −
n-Alkane C_{13}	− − −
n-Alkane C_{14}	− − −
n-Alkane C_{15}	− − −
n-Alkane C_{16}	− − −
n-Alkane C_{17}	− − −
n-Alkane C_{18}	− + −
n-Alkane C_{19}	− − −
n-Alkane C_{20}	+ + −
n-Alkane C_{21}	+ − +
n-Alkane C_{22}	+ + −
n-Alkane C_{23}	− − −
n-Alkane C_{24}	− − −
n-Alkane C_{25}	− − −
n-Alkane C_{26}	− − +
n-Alkane C_{27}	− − −
n-Alkane C_{28}	− − +
n-Alkane C_{29}	− − +
n-Alkane C_{30}	+ − +
n-Alkane C_{31}	+ − +
Benzene, alkylbenzenes	
Benzene	− + +
Toluene	− − +
Styrene	− + −
Methyl styrene	+ + −
Dimethylstyrene	− + −
Trimethyl styrene	− + −
Benzaldehyde	− + +
Acetophenone	+ + +
Xylene I	− + −
Xylene II	+ + −
Xylene III	+ + −
Propylbenzene	+ + +
C_3-Alkylbenzene I	+ − −
C_3-Alkylbenzene II	+ − −
C_3-Alkylbenzene III	+ − +
C_3-Alkylbenzene IV	− − −
C_4-Alkylbenzene I	+ + −
C_4-Alkylbenzene II	− + +
C_4-Alkylbenzene III	+ + −
C_4-Alkylbenzene IV	+ − +
C_4-Alkylbenzene V	+ + −
C_5-Alkylbenzene I	+ + −
C_5-Alkylbenzene II	+ + −
C_6-Alkylbenzene	+ + +
C_7-Alkylbenzene	+ + −
C_8-Alkylbenzene	+ + −
C_9-Alkylbenzene	+ + −
C_{10}-Alkylbenzene	+ + +
C_{11}-Alkylbenzene	+ + +
C_{12}-Alkylbenzene	+ − +
C_{13}-Alkylbenzene	− − +
C_{14}-Alkylbenzene	− − +
C_{15}-Alkylbenzene	+ − −
C_{16}-Alkylbenzene	− − −
C_{17}-Alkylbenzene	− − −
C_{18}-Alkylbenzene	− − −
C_{19}-Alkylbenzene	− − −
C_{20}-Alkylbenzene	− − −
C_{21}-Alkylbenzene	− − −
C_{22}-Alkylbenzene	− + +
C_{23}-Alkylbenzene	− − +
C_{24}-Alkylbenzene	− − +
C_{25}-Alkylbenzene	− + +
Fatty acids	
Decanoic acid	+ + +
Dodecanoic acid	+ + +
Tetradecanoic acid	+ + +
Hexadecanoic acid	+ + +
Octadecanoic acid	+ + −
Eicosanoic acid	+ − +
Docosanoic acid	+ − +
Methoxyphenols	
Guaiacol	+ + +

Compound	Environmental proxy value[a]
Methylguaiacol	+ − +
Ethylguaiacol	+ + +
Vinylguaiacol	+ − +
Propenylguaiacol	+ + +
Acetoguaiacone	+ − +
Syringol	+ + −
Methylsyringol	+ + +
Ethylsyringol	+ + +
Vinylsyringol	+ + +
Propenylsyringol	+ + +
Acetosyringone	+ + −
N-compounds	
Indole	+ + −
1*H*-Indole, 7-methyl-	+ + +
Pyridine	− + +
Methylpyridine I	− + +
Methylpyridine II	+ + +
Dimethylpyridine	+ + −
Pyrrole	− + −
1*H*-Pyrrole, 1-methyl-	+ + +
1*H*-Pyrrole, 3-methyl-	+ + −
Pyrazole	− − +
Benzonitrile	− + +
Methylbenzonitrile I	− + +
Quinoline	− + −
1*H*-Benzimidazole, 1-ethyl-	− − −
Olefins	
n-Alkene C_9	− − −
n-Alkene C_{10}	+ − −
n-Alkene C_{11}	− − −
n-Alkene C_{12}	− − −
n-Alkene C_{13}	− − +
n-Alkene C_{14}	− − −
n-Alkene C_{15}	− − −
n-Alkene C_{16}	− − −
n-Alkene C_{17}	− − +
n-Alkene C_{18}	+ + −
n-Alkene C_{19}	− − −
n-Alkene C_{20}	− + −
n-Alkene C_{21}	− − +
n-Alkene C_{22}	− − −

Compound	Environmental proxy value[a]
n-Alkene C_{23}	+ − −
n-Alkene C_{24}	− − −
n-Alkene C_{25}	− − −
n-Alkene C_{26}	− + −
n-Alkene C_{27}	− + −
n-Alkene C_{28}	− − +
Polycyclic hydrocarbons	
Azulene	+ + −
Naphthalene	+ + −
Methylnaphthalene I	+ + +
Methylnaphthalene II	+ + +
C_2-Alkylnaphthalene I	− − +
C_2-Alkylnaphthalene II	− − +
C_2-Alkylnaphthalene III	+ + +
C_2-Alkylnaphthalene IV	− + −
C_2-Alkylnaphthalene V	− + +
C_2-Alkylnaphthalene VI	− + −
C_3-Alkylnaphthalene I	− + −
C_3-Alkylnaphthalene II	− + −
C_3-Alkylnaphthalene III	+ + −
C_3-Alkylnaphthalene IV	+ + −
C_3-Alkylnaphthalene V	− + +
C_4-Alkylnaphthalene I	+ + −
C_4-Alkylnaphthalene II	+ + −
C_4-Alkylnaphthalene III	+ + +
C_4-Alkylnaphthalene IV	+ − −
Fluorene	+ + −
Methylfluorene	+ + −
Pyrene	− + −
Phenanthrene	+ + −
Anthracene	− + −
Methylphenanthrene I	− − −
Methylphenanthrene II	− − −
Methylphenanthrene III	− − +
Methylphenanthrene IV	− − −
Methylphenanthrene V	+ + −
C_2-Alkylnaphthalene I	+ + +
C_2-Alkylnaphthalene II	− − −
C_2-Alkylnaphthalene III	− + +
Biphenyl	+ + +
Diphenylmethane	+ + +

Compound	Environmental proxy value[a]
Biphenyl, 3-methyl	+ + +
Biphenyl, 4-methyl	+ + −
Biphenyldiol	− + −
Indane	+ + −
Methylindane	+ + −
Indene	− − +
Methylindene I	− − +
Methylindene II	− − +
C_2-Alkylindene I	− − −
C_2-Alkylindene II	− + −
C_2-Alkylindene III	− − −
C_2-Alkylindene IV	− − −
C_2-Alkylindene V	− − +
Inden-1-one,2,3-dihydro	− − +
Phenols	
Phenol	+ + +
Cresol I	− + −
Cresol II	− + +
Ethylphenol	− − −
4-Vinylphenol	+ + +
Carbohydrate derivatives	
2-Furancarboxaldehyde	− + −
Acetylfuran	− − −
2-Furancarboxaldehyde, 5-methyl-	+ + −
Benzofuran	+ + −
Methylbenzofuran I	− + +
Methylbenzofuran II	+ + +
Dimethylbenzofuran	+ + −
Maltol	+ + +
Levoglucosan	+ + +
Steroids	
7-Dehydrodiosgenin	− − −
Ergosta-4,6,8(14),22-tetraen-3-one	+ − −
Ergosta-5,7,9-trien-3-one	− − −
Sitosterol	− + −
Hydroxydioxocholestenyl acetate	− − −
Stigmastan-3,5-diene	+ − −
Stigmasta-3,5-dien-7-one I	+ + −
Stigmast-4-en-3-one	+ + −
Stigmasta-3,5-dien-7-one II	+ + −
Terpenoids	
Lup-20(29)-en-3-one	− + −
Friedelan-3-one	+ + −
Neoleana-3(5),12-diene	− − +

[a]: Key to its value as marker compound for the different formation factors: climate, vegetation, rock, respectively (Student's *t* test between soil groups ($P < 0.1$)).

Roman numbers indicate different isomers.

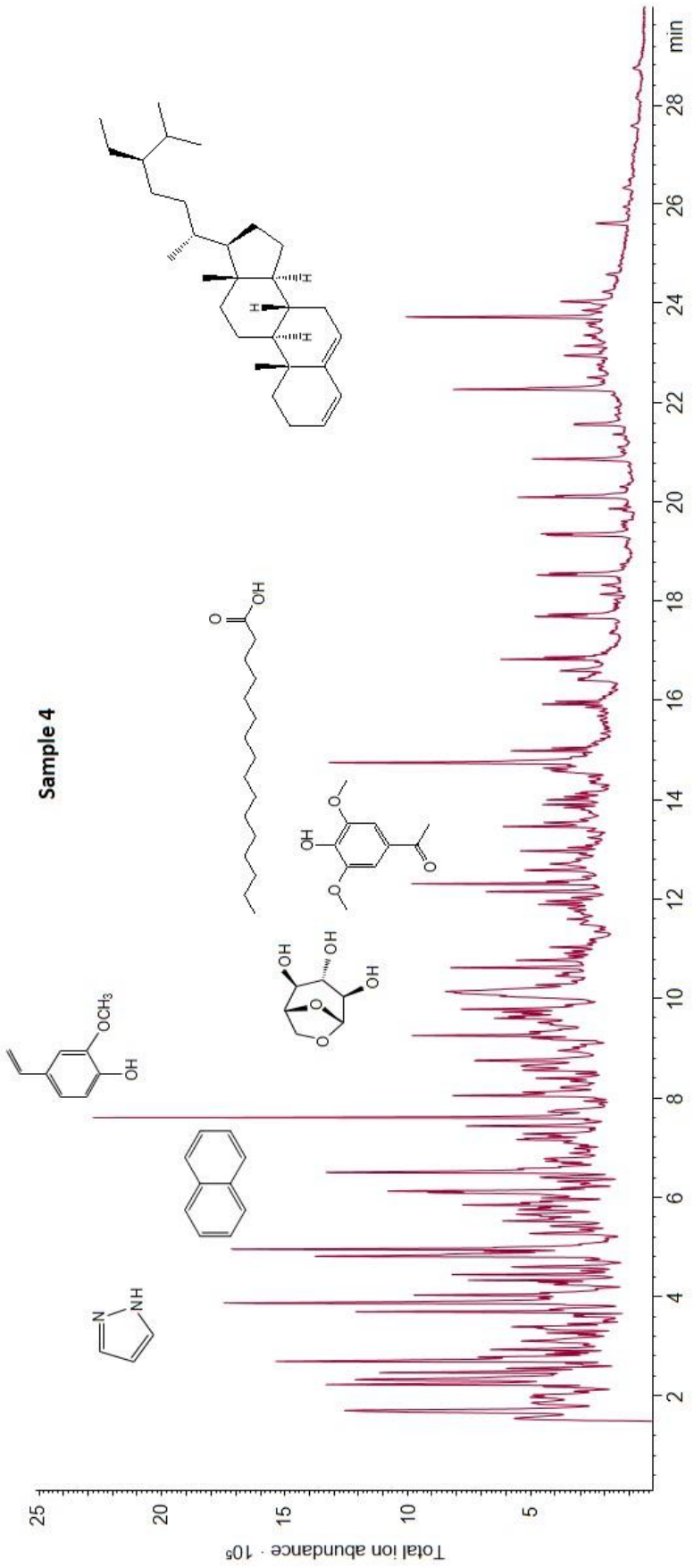


Figure 4. Example of chromatogram obtained by analytical pyrolysis of whole soil sample.

Table 11. Number of compounds identified by FTICR-MS from humic acids of the soils studied

Sample No.	Total com-pounds	Identified com-pounds	Identified com-pounds (%)	CHO	CHON	CHO NS	CHO P	CHO PN	CHO PNS	CHO PS	CHO S
1	2546	2124	83	1639	403	29	11	20	5	2	15
2	2951	2490	84	1745	662	23	1	40	2	3	14
3	2396	2021	84	1477	474	26	3	20	0	0	21
4	2836	2419	85	1946	374	29	14	20	4	1	31
5	2758	2388	87	1884	419	19	14	24	3	0	25
6	1449	1149	79	808	302	16	8	3	5	1	6
7	3009	2518	84	2136	304	24	3	32	2	2	15
8	2968	2366	80	1978	239	39	16	78	1	1	14
9	3701	2915	79	2388	350	46	13	76	1	1	40
10	4476	3487	78	2849	478	31	6	91	1	4	27
11	4001	3076	77	2471	461	51	7	62	4	5	15
12	2814	2277	81	1819	312	47	9	58	1	0	31
13	3318	2639	80	2107	402	35	2	62	1	9	21
14	2642	2099	79	1648	345	22	5	37	0	5	37
15	3366	2710	81	2357	222	26	10	45	5	4	41
16	2347	1814	77	1619	63	32	14	68	2	0	16
17	2742	2116	77	1664	297	36	10	52	1	0	56
18	1745	1309	75	1048	163	27	14	28	1	1	27
19	4005	3130	78	2496	469	49	5	73	5	2	31
20	3008	2317	77	1928	262	41	17	41	2	2	24
21	2117	1723	81	1422	225	23	10	23	1	0	19
22	3426	2881	84	2360	412	31	3	60	6	2	7
23	3125	2594	83	1790	722	24	2	31	10	1	14
24	3609	3148	87	2551	529	22	2	23	5	0	16
25	3499	3004	86	2255	694	15	0	25	5	0	10
26	2222	1616	73	885	671	24	2	10	8	0	16
27	2248	2010	89	1490	473	9	7	9	2	1	19
28	1829	1555	85	1212	311	13	6	1	1	0	11
29	3330	2889	87	2557	250	22	1	26	4	3	26
30	3477	2954	85	2673	210	23	2	22	4	2	18
31	3696	3128	85	2861	187	15	2	34	6	2	21
32	3345	2857	85	2316	465	23	9	13	11	0	20
33	3584	3078	86	2708	263	29	6	40	4	1	27
34	3378	2776	82	2395	291	25	4	25	3	1	32
35	1684	1409	84	1304	48	11	8	6	3	0	29

Table 12. Total abundances (%) of the different families of compounds identified by analytical pyrolysis of whole soil samples

Sample No.	Phenols	Alkylben-zenes	Methoxy-phenols	Alkanes	N-com-pounds	Polycyclic aromatic hydrocar-bons	Fatty acids	Olefins	Carbo-hydrate deriva-tives	Steroids
1	8.5	21.7	8.7	10.0	19.0	9.5	0.7	12.5	8.1	1.4
2	8.1	28.0	4.9	5.8	14.9	13.2	0.7	13.6	9.4	1.2
3	11.8	19.7	21.3	5.4	14.8	8.7	0.6	9.8	5.9	2.0
4	17.7	11.1	22.6	8.6	11.1	5.2	1.5	11.3	9.4	1.4
5	13.0	18.0	13.1	6.3	12.8	7.0	1.5	12.9	13.7	1.7
7	5.2	21.4	4.4	10.5	17.1	6.9	0.0	20.8	11.5	2.1
8	12.5	17.4	16.7	7.2	12.3	9.5	1.1	15.6	7.3	0.3
9	7.3	21.6	5.1	13.3	14.6	10.1	0.5	17.2	9.7	0.6
10	10.8	18.5	27.8	3.1	6.9	8.2	3.2	8.6	11.9	0.9
11	14.2	18.0	26.8	2.2	8.5	5.6	1.5	5.9	15.4	1.9
12	10.4	16.1	17.7	6.3	13.2	8.5	1.3	15.8	8.2	2.5
13	10.6	19.9	16.8	7.0	10.2	8.6	3.2	12.4	7.4	3.9
14	13.2	18.2	23.7	4.6	9.4	8.3	2.3	8.5	11.0	0.8
15	12.2	14.6	23.8	5.5	7.7	6.6	3.5	10.1	10.1	5.8
16	6.9	20.9	11.8	6.8	12.6	9.1	0.4	13.8	12.1	5.7
17	6.4	26.9	3.5	5.3	21.7	9.4	0.0	10.8	12.5	3.6
18	13.3	22.3	16.5	4.7	17.1	11.0	0.5	7.3	6.7	0.7
19	16.4	24.8	4.5	0.1	11.9	17.1	0.6	12.6	11.8	0.2
20	9.8	26.2	2.0	4.4	22.3	10.6	0.0	8.4	16.1	0.1
21	7.7	28.2	2.5	5.4	19.5	11.2	0.0	11.3	12.4	1.6
22	5.8	24.9	1.5	17.0	12.3	12.2	0.3	15.6	9.3	1.0
23	10.3	19.7	11.8	5.6	8.6	8.3	1.8	13.4	18.7	1.8
24	12.7	21.0	11.0	6.1	8.2	9.0	1.8	12.7	14.8	2.8
25	11.2	26.0	8.2	3.6	12.5	10.7	1.4	7.6	17.9	0.9
27	5.1	46.4	0.1	0.8	18.6	18.5	0.0	3.3	5.7	1.5
28	16.9	28.8	1.2	5.7	21.1	16.5	0.0	4.6	5.1	0.1
29	11.1	21.9	21.0	3.8	11.0	9.4	1.8	10.1	9.4	0.5
30	10.4	24.3	9.7	6.3	13.5	8.0	0.9	12.0	14.1	0.9
31	14.6	17.8	22.6	5.3	8.2	5.8	2.4	8.7	13.8	0.7
32	8.4	28.4	2.6	5.9	20.1	10.2	0.2	11.6	11.5	1.1
33	8.7	19.6	12.1	12.5	11.5	7.5	1.0	14.3	11.5	1.2
34	13.0	15.2	24.6	6.1	8.6	6.8	4.7	8.9	9.1	3.1
35	13.8	17.9	18.3	6.2	14.3	8.2	2.0	9.7	8.9	0.8

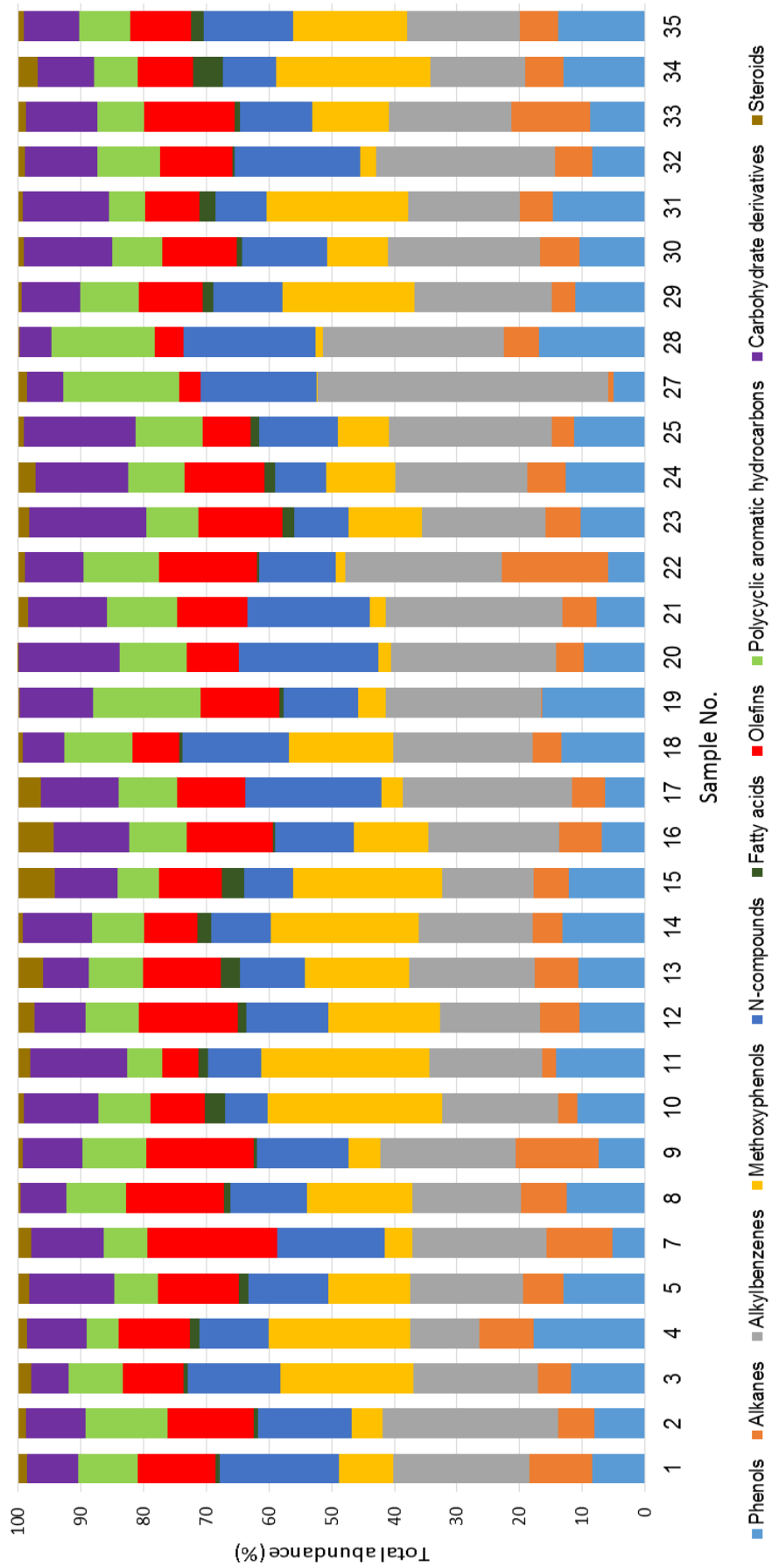


Figure 5. Proportions of the families of compounds identified by Py-GC/MS. Results referred to Table 12.

Table 13. Pyrolytic compounds showing the greatest variability in its proportions between soils at opposite quartiles (percentages represent the variation between quartiles referred to Q1 values) of the distribution of the Emberger's index (>50%), also differing significantly ($P < 0.05$) in their proportions between soils developed under wet (positive values) and arid (negative values) climates

Compound	%
Propenylsyringol	-874
Syringol	-747
Methylsyringol	-702
Vinylsyringol	-554
Ethylphenanthrene I	-476
Propenylguaiacol	-410
Acetosyringone	-323
Methylguaiacol	-210
Ethylsyringol	-192
Guaiacol	-183
Ethylguaiacol	-182
Vinylguaiacol	-175
Tetradecanoic acid	-145
1*H*-Indole, 7-methyl-	-130
Maltol	-101
C_2-Alkylphenanthrene II	89
Acetoguaiacone	-86
C_{25}-Alkylbenzene	73
C_4-Alkylnaphthalene I	70
C_4-Alkylnaphthalene II	60
C_4-Alkylnaphthalene IV	58
Indole	-57
Pyrene	56
Fluorene	53

Table 14. Compounds showing the highest values of variable importance for prediction (VIP) in the partial least squares (PLS) models to predict the Emberger's index (*Q)* from pyrolytic data from whole soil samples

Compound	VIP values
Levoglucosan	5.344
2-Furancarboxaldehyde	4.671
Phenol	4.120
2-Furancarboxaldehyde, 5-methyl	2.708
Biphenyldiol	2.658
Vinylguaiacol	2.645
***n*-Alkene C_{18}**	2.413
Benzene	2.197
4-Vinylphenol	1.954
***n*-Alkane C_{20}**	1.901
Toluene	1.798
***n*-Alkane C_{23}**	1.796
Benzonitrile	1.720
1*H*-Pyrrole, 1-methyl-	1.596
***n*-Alkene C_{20}**	1.595
Guaiacol	1.475
***n*-Alkane C_{21}**	1.385
Propenylsyringol	1.382
C_3-Alkylbenzene I	1.353
***n*-Alkane C_{27}**	1.323

Anexo II:

Artículos publicados

Journal of Chromatography A, 1508 (2017) 130–137

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

The diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of soil carbon storage☆

Marco A. Jiménez-González[a,*], Ana M. Álvarez[b], Pilar Carral[b], Francisco J. González-Vila[c], Gonzalo Almendros[a]

[a] Museo Nacional de Ciencias Naturales (MNCN-CSIC), c/Serrano 115-B, 28006 Madrid, Spain
[b] Universidad Autónoma de Madrid (UAM), c/Francisco Tomás y Valiente 7, 28049 Madrid, Spain
[c] Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS-CSIC), Av. Reina Mercedes 10, 41012 Sevilla, Spain

ARTICLE INFO

Article history:
Received 13 April 2017
Received in revised form 29 May 2017
Accepted 30 May 2017
Available online 3 June 2017

Keywords:
Soil organic matter
Analytical pyrolysis
Methoxyphenols
Diversity index
Soil carbon storage

ABSTRACT

The variable extent to which environmental factors are involved in soil carbon storage is currently a subject of controversy. In fact, justifying why some soils accumulate more organic matter than others is not trivial. Some abiotic factors such as organo-mineral associations have classically been invoked as the main drivers for soil C stabilization. However, in this research indirect evidences based on correlations between soil C storage and compositional descriptors of the soil organic matter are presented. It is assumed that the intrinsic structure of soil organic matter should have a bearing in the soil carbon storage. This is examined here by focusing on the methoxyphenols released by direct pyrolysis from a wide variety of topsoil samples from continental Mediterranean ecosystems from Spain with different properties and carbon content. Methoxyphenols are typical signature compounds presumptively informing on the occurrence and degree of alteration of lignin in soils. The methoxyphenol assemblages (12 major guaiacyl- and syringyl-type compounds) were analyzed by pyrolysis-gas chromatography-mass spectrometry. The Shannon-Wiener diversity index was chosen to describe the complexity of this phenolic signature. A series of exploratory statistical analyses (simple regression, partial least squares regression, multidimensional scaling) were applied to analyze the relationships existing between chemical and spectroscopic characteristics and the carbon content in the soils. These treatments coincided in pointing out that significant correlations exist between the progressive molecular diversity of the methoxyphenol assemblages and the concentration of organic carbon stored in the corresponding soils. This potential of the diversity in the phenolic signature as a surrogate index of the carbon storage in soils is tentatively interpreted as the accumulation of plant macromolecules altered into microbially reworked structures not readily recognized by soil enzymes. From a quantitative viewpoint, the partial least squares regression models exclusively based on total abundances of the 12 major methoxyphenols were especially successful in forecasting soil carbon storage.

1. Introduction

Despite the considerable progress achieved during the few past years in the knowledge of the factors presumptively involved in soil carbon storage, additional research is required to establish the variability of these factors in the space and time [1–3]. Such knowledge is crucial to progress in the development of the scientific bases of Earth's biogeochemical cycle and the impact of the invoked global change. Most studies center their attention on the organo-mineral interactions [3–5], whereas comparatively fewer researches have considered the possible role of the intrinsic macromolecular complexity of the soil organic matter (SOM) on the soil carbon storage. The lack of statistically-validated information of the variable role of the factors responsible of SOM stabilization in different ecosystems encourages the interest of further exploratory research looking for useful biogeochemical proxies, informing on the extent to different non-excluding factors may be active in the soil carbon storage process.

The SOM fraction presents heterogeneous structure in part derived from microbial transformation of plant biomacromolecules. Assuming the large variability of the chemical composition of the SOM, most research has focused on the analysis

☆ Selected paper from the XVI Scientific Meeting of the Spanish Society of Chromatography and Related Techniques (SECyTA 2016), 2–4 November 2016, Seville, Spain.

* Corresponding author.
E-mail address: majimenez@mncn.csic.es (M.A. Jiménez-González).

http://dx.doi.org/10.1016/j.chroma.2017.05.068
0021-9673/

of specific molecular features providing environmental information on the mechanisms and processes involved in the formation of the SOM fractions [6], which should be closely related to the performance of soil C storage. In particular, analytical tools such as pyrolysis-gas chromatography-mass spectrometry (Py-GC/MS) have lead to substantial progress not only in fingerprinting SOM structure in terms of their origin, but also in releasing molecules with high diagnostic value as biomarker compounds. In fact, recent research on the methoxyphenol composition of the fraction of humic acid (HA) in soil [7] have suggested compositional pyrolytic descriptors informative about C stabilization processes in local scenarios.

Methoxyphenols have an origin from plant biomacromolecules mainly lignin, and this family of compounds represents typical pyrolysis products from soil organic matter [8]. Unlike other plant or microbial-derived biomacromolecules which turn upon pyrolysis into secondary diagnostic products (e.g., carbohydrates leading to cyclic ketones or furans), in the case of lignins the structure of the original building blocks (guaiacyl (G), syringyl (S) and *p*-OH phenyl building blocks) are preserved to large extent.

The vegetation type in the different soils is important in the resulting methoxyphenol signature. In fact, classical studies have established the differences existing between the lignin from gymnosperms, angiosperms or grass plants in terms of the phenolic composition [9]. Major methoxyphenols released by pyrolysis from soil organic matter are *4-H-*, *methyl-*, *ethyl-*, *vinyl-*, *propenyl-* and *aceto-* derivatives of guaiacol and syringol [9] although it is well known that the monomer composition of lignins highly varies depending on the botanical source [10]. Syringol derivatives are practically lacking in pyrolysates from soils under gymnosperm plants while both S and G units are present in angiosperm plants.

On the other hand, the relative amount of G or S-type structures is very important in the biodegradability of lignin. The presence of the second methoxyl group at S-lignin units is associated to a more linear and less condensed structure. For this reason, S-lignins confer resistance with flexibility to plants, which can be important to herbaceous plants or plant organs that must grow quickly [11]. Accordingly, the use of the S/G ratio as an index for lignin degradation is based on the idea that G-type lignins are more resistant to biodegradation due to its condensed structure.

In the present research, we study the methoxyphenol signature in the SOM, in a wide variety of continental Mediterranean soils from Spain, in order to explore the presumable relationships between its composition, defined as total abundances of the different types and the resulting molecular diversity, and the performance of the soil carbon storage in the corresponding soils as reflected by multivariate chemometric models.

2. Material and methods

2.1. Soils studied

Up to 35 topsoil samples (0–10 cm) were collected from Spanish soils with contrasted values of carbon content, geological substrate, vegetation and microclimate. Three sampling points were taken in each sampling area (10 m^2). After removing the litter layer, soil materials from three sampling points were pooled to prepare the composite soil sample. Then these composites were air-dried and homogenized by sieving (fine earth, <2 mm).

2.2. Physical and chemical properties of the soils

The texture was determined following the Bouyoucos method [12]. Soil pH was measured in soil-water suspension (1:2.5), using a pH-meter pH7 (XS Instruments). A similar procedure was followed for the electrical conductivity (EC) but using soil-water slurry (1:5) and a COND7 (XS instruments). The total exchange capacity (TEC) was measured with ammonium acetate according to [13]. The soil organic carbon (SOC) was measured by the Walkley and Black wet oxidation method [14,15]. Total N was determined by micro-Kjeldahl digestion [15]. Soils classification (Table 1) was carried out according to FAO (2014).

The main chemical and spectroscopic variables of the SOM from the samples were determined aiming to identify descriptors showing correlation with SOM quantity or its quality. After removing of free organic matter (FOM) by flotation with H_3PO_4, the extractable soil organic matter was isolated with standard method using 0.1 *M* sodium pyrophosphate and sodium hydroxide. HA and fulvic acid (FA) were separated by precipitation with 6 *M* HCl [16]. For optical density measurements, solutions of humic acid in 0.1 *M* NaOH were prepared with a concentration of 0.2 mg cm^{-3} [17] to record the visible spectra and the absorbance at 465 nm (E4). The solid-state ^{13}C NMR spectra of HA were obtained in a Bruker Avance III HD 400 MHz instrument operating at a frequency of 100.64 MHz and using ZrO_2 rotors of 4 mm OD with Kel-F caps. The ^{13}C chemical shifts were calibrated relative to tetramethylsilane (0 ppm). The spectra were quantified by subdividing them into the following chemical shift regions: alkyl + α-amino C (0–45 ppm); *N*-alkyl + methoxyl C (45–60 ppm); *O*-alkyl C (60–110 ppm); aromatic C (110–160 ppm); carbonyl C (mainly carboxyl + amide 160–220 ppm). The ^{13}C intensity distribution was determined by integrating signal intensity over the above-mentioned chemical shift regions using the MestreNova 10 software.

2.3. Analytical pyrolysis

The SOM composition (whole soil samples) was studied by Py-GC/MS, which is a technique that displays large potential to assess the relative proportions of the different soil organic components [18]. To study the soil samples by analytical pyrolysis, subsamples of 2 g were reduced to fine powder (<0.01 mm) with a planetary ball mill with agate grinding bowls.

The pyrolysis was carried out with a double shot pyrolyser PY-2020iD (Frontier Lab Ltd., Fukushima, Japan) coupled to a GC/MS system Agilent 6890 equipped with a phenylmethylsiloxane column (Agilent HP-5MS 5%). Pyrolysis was conducted at 500 °C for 1 min. The carrier gas was helium at a flow rate of 1 mL min^{-1}. The GC oven temperature was held at 50 °C for 1 min, then increased to 100 °C at 30 °C min^{-1}, from 100 °C to 300 °C at 10 °C min^{-1}, and isothermal at 300 °C for 10 min. The detector consisted of an Agilent 5973 quadrupole mass spectrometer, and mass spectra were acquired at 70 eV ionizing energy. Additional single ion chromatographic traces were extracted for the different compounds in order to quantify the series of homologues or the individual compounds as total abundances. The NIST and Wiley libraries were used to confirm the identity of the 12 signature methoxyphenols: guaiacol (G), methylguaiacol (MG), ethylguaiacol (EG), vinylguaiacol (VG), propenylguaiacol (PG), acetylguaiacol (AG), syringol (S), methylsyringol (MS), ethylsyringol (ES), vinylsyringol (VS), propenylsyringol (PS) and acetylsyringol (AS).

2.4. Data analysis

Several multivariate data treatments were used to explain the origin of the variability of the concentration of SOM in the soils (dependent variable) as a function of the composition of the major pyrolysis products (independent variables). We applied the Shannon-Wiener diversity index (*H'*), classically used to characterize species diversity in a community. In this case, the 12 methoxyphenols were selected as 'species' and their total abun-

Table 1
Localization, general characteristics of soils and vegetation.

Sample No.	Geographic coordinates	Soil type (FAO)	Texture	Vegetation	Vegetation class
1	40°33′18"N 4°8′41"W	Dystric Cambisol (Humic)	Sandy loam	*Quercus pyrenaica*	Angiosperm
2	41°7′49"N 3°34′18"W	Haplic Umbrisol (Hyperhumic)	Sandy loam	*Pinus sylvestris*	Gymnosperm
3	40°23′43"N 3°16′51"W	Calcaric Cambisol (Humic)	Silt loam	*Quercus ilex*	Angiosperm
4	40°53′18"N 3°34′16"W	Gleyic Cambisol (Humic)	Sandy loam	*Fraxinus angustifolia*	Angiosperm
5	40°53′22"N 3°34′15"W	Dystric Cambisol (Humic)	Sandy loam	Pastureland	Grass
6	42°36′43"N 3°11′56"W	Calcic Chernozem (Pachic)	Sandy clay loam	Poaceae	Grass
7	40°44′1"N 3°42′22′W	Dystric Cambisol (Ochric)	Sandy loam	*Quercus rotundifolia*	Angiosperm
8	40°47′55"N 2°57′3"W	Leptic Kastanozems (Hyperhumic)	Clay loam	*Quercus rotundifolia*	Angiosperm
9	41°14′53"N 3°24′42"W	Leptic Podzol (Arenic)	Sandy loam	*Fagus sylvatica*	Angiosperm
10	40°44′53"N 3°48′18"W	Dystric Cambisol (Humic)	Sandy loam	*Pinus sylvestris*	Gymnosperm
11	40°21′30"N 3°56′7"W	Dystric Cambisol (Humic)	Loamy sand	*Pinus pinea*	Gymnosperm
12	40°33′32"N 3°43′6"W	Dystric Cambisol (Loamic)	Loamy sand	*Quercus rotundifolia*	Angiosperm
13	40°54′9"N 3°28′58"W	Eutric Cambisol (Humic)	Sandy loam	*Juniperus oxycedrus*	Gymnosperm
14	40°52′56"N 3°34′32"W	Eutric Cambisol (Humic)	Sandy loam	*Juniperus oxycedrus*	Gymnosperm
15	40°58′45"N 3°37′6"W	Dystric Cambisol (Humic)	Sandy loam	*Pinus pinaster*	Gymnosperm
16	40°54′27"N 3°53′46"W	Dystric Cambisol (Humic)	Sandy clay loam	*Quercus pyrenaica*	Angiosperm
17	40°51′41"N 3°44′4"W	Dystric Cambisol (Colluvic)	Sandy loam	*Pinus sylvestris*	Gymnosperm
18	40°45′35"N 3°41′9"W	Leptic Cambisol (Humic)	Loam	*Quercus ilex*	Angiosperm
19	43°15′34"N 2°51′19"W	Leptic Umbrisol (Loamic)	Loam	Pastureland	Grass
20	43°4′29"N 2°35′48"W	Haplic Luvisol (Humic)	Silty clay loam	*Fagus sylvatica*	Angiosperm
21	42°34′10"N 2°38′15"W	Eutric Cambisol (Humic)	Clay loam	Pastureland	Grass
22	43°15′39"N 2°51′11"W	Haplic Umbrisol (Loamic)	Loam	*Pinus radiata*	Gymnosperm
23	42°28′54"N 8°53′39"W	Leptic Regosol (Humic)	Sandy loam	*Pinus pinaster*	Gymnosperm
24	42°36′58"N 8°38′40"W	Leptic Regosol (Humic)	Sandy loam	*Pinus pinaster*	Gymnosperm
25	43°4′52"N 8°22′06"W	Leptic Umbrisol (Hyperhumic)	Loam	*Pinus pinaster*	Gymnosperm
26	28°22′48"N 16°39′19"W	Vitric Andosol (Hyperhumic)	Sandy loam	*Laurus canariensis*	Angiosperm
27	28°26′16"N 16°29′25"W	Leptic Regosol (Arenic)	Clay loam	Coastal scrub	Angiosperm
28	28°14′14"N 16°28′21"W	Leptic Regosol (Arenic)	Sandy loam	Fallow land	Grass
29	28°9′50"N 16°38′23"W	Folic Umbrisol (Chromic)	Sandy loam	*Pinus canariensis*	Gymnosperm
30	40°13′43"N 4°29′53"W	Dystric Regosol (Arenic)	Sand	*Pinus pinea*	Gymnosperm
31	41°29′18"N 4°19′19"W	Eutric Cambisol (Humic)	Sand	*Pinus pinea*	Gymnosperm
32	40°18′34"N 4°38′46"W	Eutric Cambisol (Arenic)	Sandy loam	*Quercus rotundifolia*	Angiosperm
33	41°1′38"N 3°12′26"W	Eutric Cambisol (Humic)	Silt loam	*Quercus rotundifolia*	Angiosperm
34	40°58′54"N 3°44′13"W	Eutric Cambisol (Humic)	Sandy loam	*Juniperus thurifera*	Gymnosperm
35	40°56′37"N 3°41′41"W	Dystric Leptosol (Humic)	Loam	*Juniperus thurifera*	Gymnosperm

dances were used to calculate this index with the software Species Diversity & Richness II version 2.5.

Simple regression analysis was used to compare the *H*' biodiversity with the SOC in different soil samples and in sample subsets. Multivariate data treatments were carried out with the software Statgraphics Centurion XV, using the total abundances of 12 methoxyphenols after square root transformation as independent variables. Partial least squares regression (PLS) was applied considering the 12 methoxyphenols as independent variables and the SOC content as dependent variable. Spurious models due to overfitting were detected and discarded after repeating PLS models with fully randomized dependent variables. Discriminant analysis was also performed to check for the extent to which the methoxyphenol composition of the SOM may represent a source of proxies for the prediction of vegetation types. Finally multidimensional scaling (MDS) using the software Statistica ver. 7.1 was used for simultaneous ordination of different soil dependent and independent variables, illustrating their mutual relationships as well as with the SOC levels.

3. Results

The general properties of the soils (Table 2) illustrate the large variability in the SOC (from 17 to 157 g kg^{-1}). One out of the sources of variability is the vegetation type, which in most cases also depend of the soil use. To analyze the response to vegetation with the methoxyphenol signature, two principal groups of vegetation were considered (angiosperms, gymnosperms) and a third, comparatively small group with herbaceous species referred to as the grass group (Table 1).

The 12 methoxyphenols identified by Py-GC/MS and their diagnostic ion traces are shown in Fig. 1. The peak close to S corresponds to biphenyl which also has a characteristic ion at 154 *m*/*z*.

The *H*' diversity index calculated from the total abundance of methoxyphenols (Table 2) showed large variability within the samples. When correlating SOC with *H*' separately calculated for the S and G families (Fig. 2), a highly significant correlation was observed, i.e., the increase in the diversity (or complexity) of the methoxyphenol assemblages is systematically accompanied by an increase in the SOC in the corresponding soils with a statistically significant level (*P*-value = 0.0005 for G group and *P*-value = 0.0008 for the S group).

A series of biodiversity indexes calculated from the 12 methoxyphenols (mainly Shannon-Wiener index, Simpson index, Margalef index, equitability and Berger-Parker index) were also calculated (data not shown), leading to similar conclusions and small differences as regards the significance levels. Consequently, the Shannon's *H*' was selected as being the most frequently used, not intending further justification of the different features of biodiversity which are selectively explained for these alternative indices.

When checking whether *H*' is correlated with other pyrolytic indices informing on changes in the structure of lignin, a poor correlation with the S/G ratio ($P > 0.05$) was found, but a very significant correlation exists with the ratio C_{0-2}-MeO-Ph/C_3-MeO-Ph ($P < 0.000$). This index (methoxyphenols with less than 3 alkyl C/(PG + PS) would inform on the degradation and oxidation of the C_3 side-chain in the structural units of lignin and suberin macromolecules (Fig. 2), which can be considered as a reflection of the progressive loss with decomposition of the biogenic lignin signature.

The PLS exclusively using total abundances of methoxyphenols as independent variables and including the first four PLS

Table 2
Soil chemical and physical properties

Sample	SOC ($g kg^{-1}$)	C/N	pH	EC ($\mu S\ cm^{-1}$)	TEC ($cmol + kg^{-1}$)	Bulk density ($g\ cm^{-3}$)	FOM ($g\ C\ 100\ C\ g^{-1}$ soil)	FA ($g\ C\ 100\ C\ g^{-1}$ soil)	HA ($g\ C\ 100\ C\ g^{-1}$ soil)	E4	HA aromatic C (%)	HA *O*-alkyl C (%)	HA *N*-alkyl & methoxyl C (%)	HA alkyl C (%)	HA carbonyl C (%)	Sand ($g\ kg^{-1}$)	Silt ($g\ kg^{-1}$)	Clay ($g\ kg^{-1}$)	*H'* Guaiacyl	*H'* Syringyl	*H'* Total	C_{0-2}-MeO-Ph/C_3-MeO-Ph
1	41	11.3	5.2	488	16.4	0.96	3.6	10.4	48.0	0.78	18.9	27.4	10.8	28.4	14.6	581	261	158	1.484	1.622	1.935	12.477
2	67	14.8	3.9	385	24.4	0.89	6.4	10.5	45.9	1.14	27.8	18.7	8.4	32.7	12.5	603	320	77	1.569	1.646	1.770	13.820
3	96	15.3	7.2	1143	30.8	0.81	3.5	11.1	31.7	0.66	21.9	24.2	11.3	31.4	11.3	215	758	27	1.627	1.688	2.173	6.508
4	87	13.3	6.2	747	22.0	0.85	6.5	14.7	33.4	0.38	16.8	24.8	11.7	35.4	11.3	622	264	113	1.559	1.683	2.166	4.406
5	48	13.1	5.3	418	14.4	0.86	7.9	9.0	37.4	0.51	19.1	24.8	11.8	31.6	12.7	522	352	126	1.399	1.607	1.858	10.240
6	17	13.9	7.6	1025	15.9	1.21	8.2	3.1	26.9	1.29	32.3	14.3	9.6	30.8	13.0	502	240	258	1.582	1.456	2.202	5.530
7	18	16.0	5.2	290	4.5	1.13	5.2	16.5	37.4	0.68	18.1	21.0	9.7	41.1	10.1	768	156	76	1.563	1.690	1.912	18.450
8	87	13.3	6.8	1326	21.5	0.77	5.3	5.5	36.8	0.86	22.1	22.2	9.9	32.9	12.9	267	415	318	1.632	1.724	2.207	4.649
9	32	16.4	5.7	490	11.6	1.06	3.3	10.1	17.8	0.78	18.1	21.4	10.4	37.2	12.9	593	318	89	1.615	1.690	1.990	6.745
10	140	18.1	5.1	549	14.2	0.62	21.4	12.4	41.1	0.71	24.1	23.1	9.6	32.4	10.9	768	116	116	1.652	1.713	1.708	6.515
11	117	26.7	4.9	499	13.9	0.76	35.9	7.6	19.7	0.59	22.4	25.2	11.0	31.0	10.5	828	56	116	1.617	1.754	1.735	5.963
12	93	8.9	6.4	693	20.1	0.86	5.9	14.4	41.0	0.53	16.6	26.6	10.7	33.2	12.8	848	56	96	1.593	1.664	2.154	5.763
13	134	12.1	7.0	1356	19.8	0.56	14.0	10.0	39.5	0.51	18.1	24.6	9.9	36.0	11.5	548	276	176	1.596	1.699	1.853	7.701
14	104	18.5	6.5	786	22.2	0.67	16.3	8.2	36.4	0.61	21.0	26.3	9.4	31.0	12.2	648	216	136	1.646	1.744	1.972	5.828
15	81	16.9	6.0	552	17.1	0.75	16.0	14.8	42.7	0.48	19.2	25.2	10.2	34.8	10.6	788	116	96	1.638	1.731	1.937	6.714
16	55	18.0	5.7	309	15.9	0.89	12.7	14.0	35.5	0.55	17.4	25.7	10.2	36.1	10.7	620	180	200	1.583	1.598	1.966	7.380
17	39	13.0	5.6	186	13.8	1.09	9.4	15.0	50.1	0.80	18.9	25.3	10.5	33.9	11.5	540	290	170	1.510	1.557	1.796	13.584
18	105	17.0	7.2	767	41.9	0.68	12.1	13.0	52.1	0.51	18.6	24.7	11.8	32.3	12.7	380	410	210	1.619	1.718	2.186	7.340
19	41	15.8	4.6	381	15.8	0.90	5.0	11.3	49.3	0.95	22.6	20.1	8.9	37.0	11.4	384	374	242	1.543	1.709	1.849	13.600
20	44	13.4	5.1	491	11.3	0.92	10.6	17.7	39.8	0.61	18.2	25.7	11.2	33.8	11.1	169	537	294	1.294	1.500	1.476	23.934
21	57	13.9	6.9	645	32.8	0.82	7.0	17.5	48.7	0.66	19.1	26.5	10.3	31.3	12.8	326	291	383	1.269	1.505	1.535	25.052
22	27	17.0	4.2	324	13.3	0.95	6.9	14.2	43.7	0.95	20.6	18.6	8.8	41.5	10.5	386	403	211	1.543	1.377	1.712	17.066
23	133	31.0	3.5	485	32.4	0.65	13.9	11.1	40.0	1.18	26.4	20.2	8.8	32.0	12.7	630	210	160	1.572	1.718	1.755	12.623
24	90	20.0	3.7	361	22.6	0.79	8.4	12.6	54.4	0.76	21.7	22.3	8.9	35.0	12.0	670	220	110	1.520	1.692	1.740	15.551
25	132	18.0	4.2	527	31.7	0.54	9.0	11.8	39.7	0.76	21.9	26.0	10.3	29.7	12.1	320	420	260	1.564	1.675	1.751	12.451
26	18	22.0	6.8	514	28.3	0.98	0.3	2.0	18.8	1.60	35.1	16.8	7.9	25.5	14.8	610	290	100	1.524	1.660	2.249	4.532
27	22	12.0	7.7	850	18.3	1.06	17.5	11.1	43.7	0.67	20.4	19.2	10.9	37.0	12.5	420	300	280	1.349	1.163	1.949	30.005
28	23	14.0	7.3	784	29.1	0.99	9.4	7.5	38.5	0.72	22.6	21.7	11.3	31.9	12.5	600	230	170	1.127	1.118	1.266	34.015
29	105	27.0	6.4	475	20.2	1.01	10.7	8.0	23.6	0.55	24.1	22.5	10.1	32.9	10.4	680	140	180	1.606	1.702	1.890	8.134
30	35	20.0	5.7	327	4.9	1.11	18.5	13.4	22.6	0.66	18.4	21.4	10.5	39.5	10.3	935	27	38	1.446	1.714	1.530	17.439
31	99	25.9	5.7	561	9.3	0.64	23.5	7.1	10.2	0.58	20.8	22.4	10.2	36.8	9.8	913	30	58	1.620	1.739	1.731	5.962
32	46	16.8	5.4	376	12.8	0.94	5.0	9.0	43.3	1.13	25.0	21.3	8.7	31.5	13.6	717	149	135	1.350	1.701	1.592	40.682
33	89	23.7	6.2	759	17.2	0.64	23.2	11.6	17.7	0.64	20.3	24.1	10.3	35.4	9.8	378	488	134	1.590	1.678	2.092	5.689
34	157	21.6	6.6	1087	18.0	0.59	28.0	7.0	13.6	0.46	18.8	23.7	10.9	37.3	9.4	637	167	196	1.655	1.722	1.983	4.965
35	92	13.9	7.4	820	25.9	0.91	4.3	17.4	47.1	0.62	17.8	25.0	11.0	34.8	11.5	392	384	224	1.657	1.729	1.972	4.605

SOC: soil organic carbon; EC: electrical conductivity; TEC: total exchance capacity; FOM: free organic matter; FA: fulvic acid; HA: humic acid; E4: optical density of the HA solutions at 465 nm; *H'*: Shannon index; C_{0-2}-MeO-Ph/C_3-MeO-Ph: (G + MG + EG + VG + AG + S + MS + ES + VS + AS)/(PG + PS).

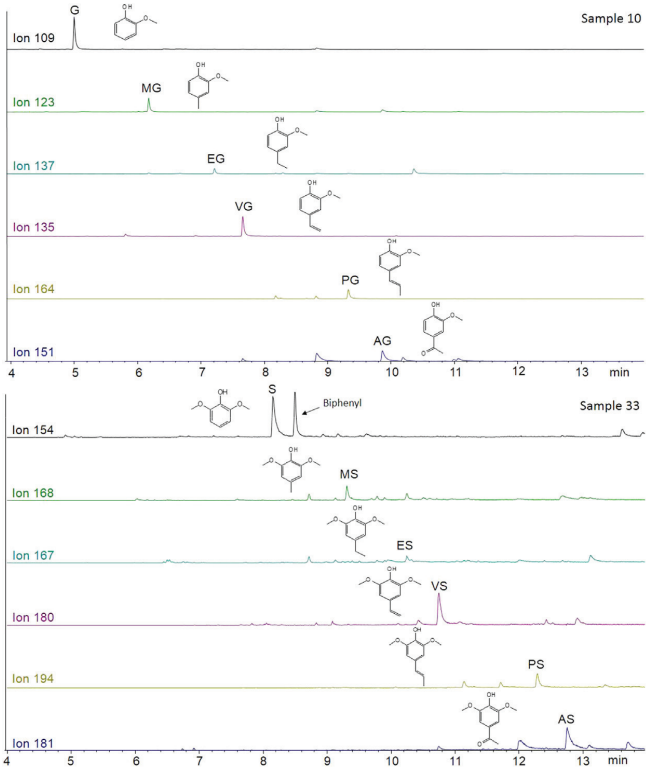


Fig. 1. Chromatographic traces for diagnostic ions of the major methoxyphenols released by analytical pyrolysis from SOM (guaiacol (G), methylguaiacol (MG), ethylguaiacol (EG), vinylguaiacol (VG), propenylguaiacol (PG), acetylguaiacol (AG), syringol (S), methylsyringol (MS), ethylsyringol (ES), vinylsyringol (VS), propenylsyringol (PS) and acetylsyringol (AS).

components led to highly significant ($P < 0.01$) forecasting models explaining the concentration of organic C in the different soils (Fig. 3). The validity of the models was confirmed by comparison with the alternative models computed from the fully randomized SOC (values of SOC assigned to the soil that not correspond with the real value in the soil) as dependent variable, which had a poor correlation ($P > 0.05$) with the total abundances of the methoxyphenols, whereas the real SOC values resulted in an excellent correlation with the methoxyphenol composition. In Fig. 3 (cross validation plot) the observed values are represented *versus* the predicted values. The standardized coefficients for the methoxyphenols, obtained by PLS, are also shown in Fig. 3.

The results obtained in the discriminant analysis show an outstanding prediction potential of the methoxyphenols as regards vegetation type (gymnosperms, angiosperms or grasses). In Fig. 4 it is possible to observe the soil samples grouped in three sharp clusters. The three groups of vegetation in the 35 samples include 97% of cases correctly classified.

The results of MDS analysis are represented as a scatter diagram in Fig. 5. In this plot the scores for the different variables are proportional to the 1-Pearson indices from the correlation matrix including dependent and independent variables. In this plot the variables represented by points more close in the plane also correspond to those with more significant correlation indices between. The 'stress' level in the diagram was 0.2169, indicating excellent reliability of the scatterplot [19] (i.e., low spatial distortion when the original *n*-dimensional space is represented in the 2D plane).

In particular, when comparing the correlations between SOC and a large number of variables informing on soil physical and chemical characteristics, or on the concentration of the main soil organic fractions, or even the amounts of the major constituents of HA analyzed by ^{13}C NMR, it is found that the *H*' biodiversity indices of the methoxyphenols are the only variables very close to the SOC, with correlation indices $P < 0.001$ (Fig. 5).

At a next level, other variables are only correlated at $P < 0.01$ with the SOC, e.g., FOM and percentage of *O*-alkyl C in HA, i.e., variables indicative of accumulation of litter-derived organic matter. Finally, other soil variables as C/N ratio and EC correlated with SOC only at $P < 0.05$ level.

4. Discussion

Recent studies have centered on factors such as physical protection of SOM by soil minerals that affect on the retention of SOM in different depths [20]. Nevertheless, in our opinion, the large variability in the soil carbon content (Table 2) in our samples probably depends on a plethora of factors, which is presumptively reflected by the complex composition of pyrolytic compounds.

In this sense, most recent literature also tend to consider soil C storage as a scale-dependant process largely influenced by local factors, avoiding ideal models based on the impact of global processes. In fact, carbon storage has been considered to be largely dependent on physicochemical factors such as clay content, nutrient status and water regime and temperature on the soil [21]. On

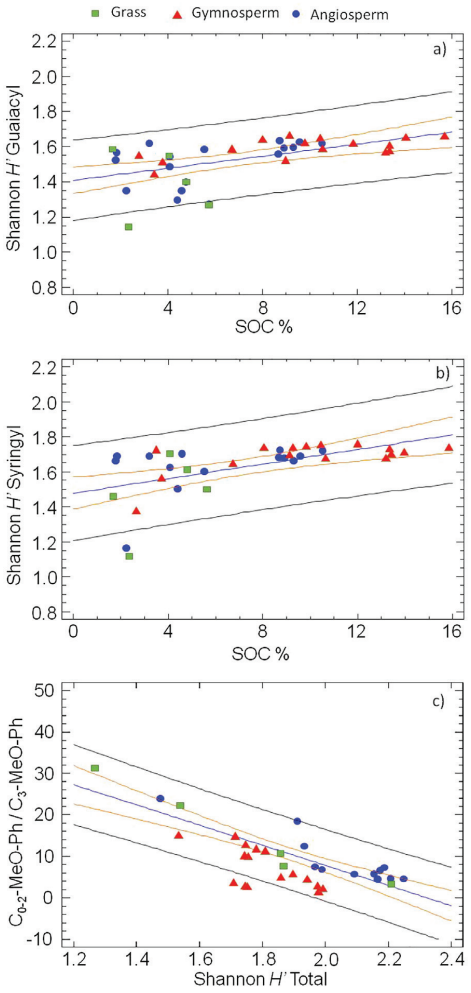


Fig. 2. a, b) Relationship between the concentration of soil organic carbon (SOC)and the Shannon's biodiversity index (*H'*) calculated from the 12 major methoxyphenols released from soils by Py-GC/MS and c) biogenic methoxyphenol signature expressed as ($C_{0\text{-}2}$-MeO-Ph/C_3-MeO-Ph), with Shannon's biodiversity index. Inner and outer lines indicated 95% confidence and prediction limits, respectively.

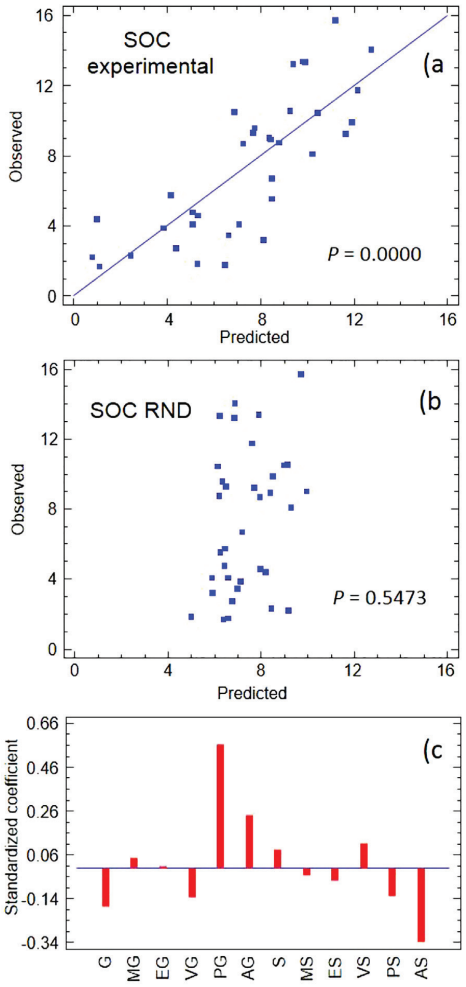


Fig. 3. Observed versus predicted values of soil organic carbon (SOC) content calculated by partial least squares regression (PLS) from a) experimental values, b) fully randomized SOC values; c) standardized coefficients of 12 methoxyphenols obtained for SOC in the PLS forecasting model.

the other hand, water loss, soil erosion, and soil mismanagement induced a great decrease of the SOC [22]. Factors associated with vegetative cover were major influences on SOC accumulation [23]. In other cases, the key role of the composition of the soil solution has also been considered crucial in determining the final SOC values [24].

Assuming that large number of factors is involved in soil C storage, it is worthy to point out that the results in this study indicated than only a limited subset of the whole pyrolytical assemblage i.e., the 12 major methoxyphenols, provided sufficient insight on the SOM concentration, its quality and some additional information on the botanical origin of the major precursors of the soil organic matter.

Apart from the studies considering C stabilization as a process highly dependent on physical interactions in soil, alternative studies have pointed out that the resistance to biodegradation of the SOM may also differ to large extent in laboratory incubation systems in which organo-mineral interactions were not responsible for the origin of the variability [6]. In such a system, multivariate models explained the CO_2 release as resulting from a series of structural variables in the HA, none of which having the crucial role. This supports the idea that, as whole, the structural variability of humic matter is an important factor in SOM resilience. In fact, the above mutual correlations between important soil variables with a presumptive role on biodegradation and humification processes, suggest that the molecular composition of the SOM is more a useful predictor of the SOC levels than the relative amounts of the more or less decomposed, particulate or colloidal SOM fractions.

The large variability in soil types and processes ought also to be taken into consideration. For instance, previous studies on agricultural soils where the vegetation was the same but the geological substrate differed to large extent [25] evidenced the expected

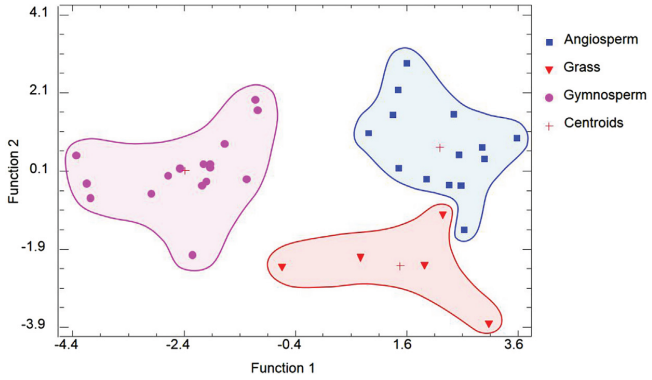


Fig. 4. Discriminant analysis using the 12 pyrolytic methoxyphenols as variables and the vegetation type (angiosperm, gymnosperm and grass) as classification factor.

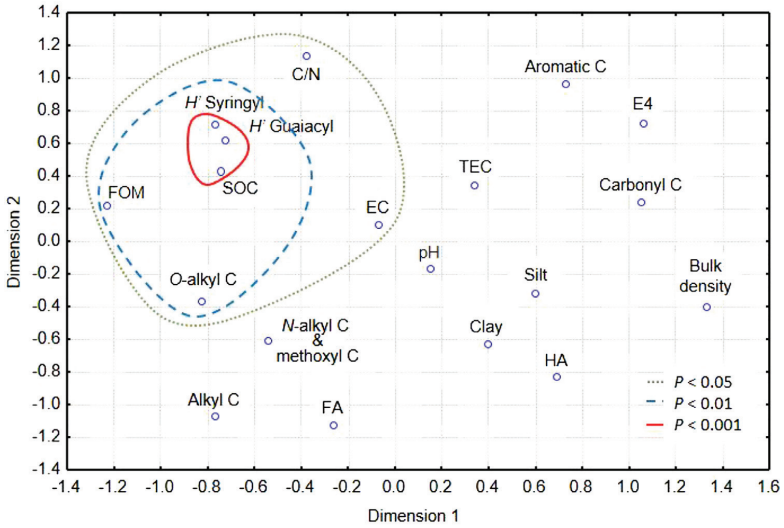


Fig. 5. Representation, using multidimensional scaling (MDS), of the relationships between different soil properties and the Shannon *H'* diversity index of pyrolytic methoxyphenols. Variables related with progressively less significant *P* values with the SOC are encircled with continuous, dashed and dotted lines, respectively.

importance of the accumulation of amorphous minerals in justifying low CO_2 release rates and, conversely, high SOM levels in the soils. In particular, the entrapment of SOM on microporous matrices under local anaerobic conditions has been recently postulated as a main process enhancing preservation of SOM of microbial origin in volcanic ash soils [26]. Nevertheless, even in this extreme scenario with volcanic minerals with recognized protective role of SOM, much of the total variance in the resulting resilience was explained by structural factors. This effect was more conspicuous when the total aromaticity—with poor forecasting potential as regards SOM levels [25]—was studied after isolating the aromaticity due to lignin constituents (i.e., methoxyphenols) from that due to condensed, alkylbenezenoid structures, including pyrogenic OM. In this study, the fact that some topsoils accumulate more organic matter than others is explained by the structure of the soil organic matter, reflected in the biogenic methoxyphenol signature which also coincides with a low degree of alteration of macromolecules in soil (mainly lignins and suberins).

5. Conclusions

The analysis of the pyrolytic methoxyphenol signature of continental Mediterranean soils suggests that a relation exists between the diversity calculated from the relative concentration of the 12 major phenols and the total SOM concentration. This finding is interpreted as C storage behave as a soil emergent property, which is reflected by the sources, complexity and interactions of the soil organic matter constituents at a molecular level. Different statistical analysis (simple regression, PLS, MDS) coincide in pointing out to a relation between C storage in soil and the methoxyphenol complexity. Apart from this, the close relationship between the vegetation type and the soil pyrolytic methoxyphenol signature was corroborated by discriminant analysis. Although no cause-to-effect relationships are postulated from the fact that the biodiversity in the methoxyphenol assemblages parallel the soil C concentration, it is clear that this fact ought to be considered in the current controversy on the origin of SOM recalcitrance, which

in our samples can be expressed as a function of pyrolytic proxies informing on the structural complexity of the SOM.

Acknowledgements

Financial support by Spanish CICyT (grant CGL2013-43845-P) is gratefully acknowledged. Marco A. Jiménez-González thanks the Spanish Ministry of Economy and Competitiveness (MINECO) for funding his pre-doctoral FPI fellowship (BES-2014-069238). The authors are grateful to anonymous reviewers for their valuable comments and suggestions on the manuscript draft.

References

[1] R.S. Swift, Sequestration of carbon by soil, Soil Sci. 166 (2001) 858–871, http://dx.doi.org/10.1097/00010694-200111000-00010.
[2] H. Yuan, T. Ge, M. Kleber, C. Chen, A.G. O'Donnell, J. Wu, Significant role for microbial autotrophy in the sequestration of soil carbon, Appl. Environ. Microbiol. 78 (2012) 2328–2336, http://dx.doi.org/10.1128/AEM, 06881-11.
[3] F. Wenting, A.F. Plante, A.K. Aufdenkampe, J. Six, Soil organic matter stability in organo-mineral complexes as a function of increasing C loading, Soil Biol. Biochem. 69 (2014) 398–405, http://dx.doi.org/10.1016/j.soilbio.2013.11.024.
[4] I. Kögel-Knabner, G. Guggenberger, M. Kleber, E. Kandeler, K. Kalbitz, S. Scheu, K. Eusterhues, P. Leinweber, Organo-mineral associations in temperate soils: integrating biology, mineralogy, and organic matter chemistry, J. Plant Nutr Soil Sci. 171 (2008) 61–82, http://dx.doi.org/10.1002/jpln.200700048.
[5] M.V. Lützow, I. Kögel-Knabner, K. Ekschmitt, E. Matzner, G. Guggenberger, B. Marschner, H. Flessa, Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions–a review, Eur. J. Soil Sci. 57 (2006) 426–445, http://dx.doi.org/10.1111/j. 1365-2389.2006.00809.
[6] Z. Hernández, G. Almendros, P. Carral, A. Álvarez, H. Knicker, J.P. Pérez Trujillo, Influence of non-crystalline minerals in the total amount, resilience and molecular composition of the organic matter in volcanic ash soils (Tenerife Island, Spain), Eur. J. Soil Sci. 63 (2012) 603–615, http://dx.doi.org/10.1111/j.1365-2389.2012.01497.x.
[7] I. Miralles, A. Piedra-Buena, G. Almendros, F.J. González-Vila, J.A. González-Pérez, Pyrolytic appraisal of the lignin in soil humic acids: assessment of its usefulness as carbon sequestration marker, J. Anal. Appl. Pyrolysis 113 (2015) 107–115, http://dx.doi.org/10.1016/j.jaap.2014.11.010.
[8] F. Martín, C. Sáiz-Jiménez, F.J. González-Vila, Pyrolysis–gas chromatography–mass spectometry of lignins, Holzforschung 33 (1979) 210–212, http://dx.doi.org/10.1515/hfsg.1979.33.6.210.
[9] P. Tinoco, G. Almendros, F.J. González-Vila, Impact of the vegetation on the lignin pyrolytic signature of soil humic acids from Mediterranean soils, J. Anal. Appl. Pyrolysis 64 (2002) 407–420, http://dx.doi.org/10.1016/S0165-2370(02)00041-4.
[10] D. Fengel, G. Wegener, Wood: Chemistry, Ultrastructure, Reactions, De Gruyter Berlin, New York, 1983.
[11] N.D. Bonawitz, C. Chapple, The genetics of lignin biosynthesis: connecting genotype to phenotype, Annu. Rev. Genet. 44 (2010) 337–363, http://dx.doi.org/10.1146/annurev-genet-102209-163508.
[12] G.J. Bouyoucos, The hydrometer as a new method for the mechanical analysis of soils, Soil Sci. 23 (1927) 343–353.
[13] A.S.R. Juo, S.A. Ayanlaja, J.A. Ogunwale, An evaluation of cation exchange capacity measurements for soils in the tropics, Commun. Soil Sci. Plant Anal. 7 (1976) 751–761, http://dx.doi.org/10.1080/00103627609366684.
[14] A. Walkley, I.A. Black, An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method, Soil Sci. 37 (1934) 29–38.
[15] D.V. Nelson, L.E. Sommers, Total carbon, organic carbon and organic matter, in: A.L. Page, R.H. Miller, D.R. Keeney (Eds.), In: Methods of Soil Analysis: Part 2, Chemical and Microbiological Properties, second ed., American Society of Agronomy, Madison WI, 1982, pp. 539–579.
[16] P. Duchaufour, F. Jacquin, Comparaison des processus d'humification dans les principaux types d'humus forestiers, Bulletin AFES 1 (1975) 29–36.
[17] M.M. Kononova, Materia orgánica del suelo: su naturaleza, propiedades y métodos de investigación, Barcelona, Oikos Tau, 1982, http://www.worldcat.org/title/materia-organica-del-suelo-su-naturaleza-propiedades-y-metodos-de-investigacion/oclc/632426606.
[18] J.M. De la Rosa, J.A. González-Pérez, R. González-Vázquez, H. Knicker, E. López-Capel, D.A.C. Manning, F.J. González-Vila, Use of pyrolysis/GC–MS combined with termal analysis to monitor C and N changes in soil organic matter from Mediterranean fire affected forest, Catena 74 (2008) 296–303, http://dx.doi.org/10.1016/j.catena.2008.03.004.
[19] J.B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika 29 (1964) 1–27, http://dx.doi.org/10.1007/BF02289565.
[20] M.W.I. Schmidt, M.S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I.A. Janssens, M. Kleber, I. Kögel-Knabner, J. Lehmann, D.A.C. Manning, P. Nannipieri, D.P. Rasse, S. Weiner, S.E. Trumbore, Persistence of soil organic matter as an ecosystem property, Nature 478 (2011) 49–56, http://dx.doi.org/10.1038/nature10386.
[21] J.M. Oades, The retention of organic matter in soils, Biogeochemistry 5 (1988) 35–70, http://dx.doi.org/10.1007/BF02180317.
[22] R. Lal, Soil carbon sequestration to mitigate climate change, Geoderma 123 (2004) 1–22, http://dx.doi.org/10.1016/j.geoderma.2004.01.032.
[23] M.L. Dillon, Explaining soil organic carbon sequestration in an urban ecosystem, 96th ESA Annual Meeting (2011). COS 193-7, Austin, Texas.
[24] C. Pan, H. Zhao, X. Zhao, H. Han, Y. Wang, J. Li, Biophysical properties as determinants for soil organic carbon and total nitrogen in grassland salinization, PLoS ONE 8 (2013), http://dx.doi.org/10.1371/journal.pone.0054827.
[25] Z. Hernández, G. Almendros, Biogeochemical factors related with organic matter degradation and C storage in agricultural volcanic ash soils, Soil Biol. Biochem 44 (2012) 130–142, http://dx.doi.org/10.1016/j.soilbio.2011.08.009.
[26] P. Buurman, F. Peterse, G. Almendros, Soil organic matter chemistry in allophanic soils: a pyrolysis-GC/MS study of a Costa Rican Andosol catena, Soil Sci. 58 (2007) 1330–1347, http://dx.doi.org/10.1111/j.1365-2389.2007.00925.x.

Biology and Fertility of Soils (2018) 54:617–629
https://doi.org/10.1007/s00374-018-1285-6

ORIGINAL PAPER

Soil carbon storage predicted from the diversity of pyrolytic alkanes

Marco A. Jiménez-González[1,2] · Ana M. Álvarez[2] · Zulimar Hernández[2] · Gonzalo Almendros[1]

Received: 19 January 2018 / Revised: 25 April 2018 / Accepted: 27 April 2018 / Published online: 7 May 2018

Abstract
Biogeochemical factors responsible of the highly variable content of soil organic matter (SOM) in the different types of soils are poorly known. In particular, the role of organo-mineral interactions has frequently been considered, but less attention has been paid to the molecular composition of the SOM. The aim of this work was to contribute to a better qualitative and quantitative assessment of the soil organic C (SOC) accumulation, using chemometric approaches that do not require the absolute knowledge of the structure and functioning of the whole system under study. For this reason, we monitored the *n*-alkanes released by analytical pyrolysis from 35 widely different Mediterranean soils. The H' Shannon diversity index was calculated to evaluate the origin and transformations of the alkane homologous series (C_9–C_{31}). A series of multivariate data treatments succeeded in showing significant relationship between the diversity of alkanes and the SOC concentration, and additional indicators of SOM quality were also used. All statistical analyses pointed out the significant correlation ($P < 0.01$) between the H' diversity of the pyrolytic alkanes and the amount of SOC. In particular, a significant relationship between SOC levels and the percentage of long-chain alkanes was found, whereas the percentage of short-chain alkanes was correlated with specific descriptors of SOM quality. Finally, the partial least squares (PLS) predicted the SOC content from the alkane patterns.

Keywords Soil alkanes · Analytical pyrolysis · Shannon biodiversity index · Soil carbon storage · Soil organic matter

Introduction

The biogeochemical factors involved in the structural transformations of soil organic matter (SOM) are highly variable in the spatial and temporal scales, and for this reason, soils have different concentrations of soil organic C (SOC). Therefore, accurate monitoring of soil C storage processes is crucial for a quantitative assessment of the global C cycle, because the total amount of organic C stored in soil not only is an important C pool excluded from the atmospheric circulation (Lal 2004) but also depends on the stabilized, humic-type organic matter (Kögel-Knabner et al. 1988). For these reasons, in recent decades, much research has focused on modeling the C storage of different soils. Generally, the stabilization of SOM has been simulated by simple models mainly considering SOC preservation through organo-mineral interactions (Schmidt et al. 2011) and met the molecular structure of SOM and the specific enzyme processes. However, in vitro incubation experiments have showed a large variability in the rates of transformation of plant and microbial residues into stabilized humic-type material (Almendros and Dorado 1999; Pendall and King 2007; Requena et al. 1996). Other research showed how microbial and fungal community to large extent depends on the composition of different SOM fractions (Montiel-Rozas et al. 2017; Tian et al. 2017).

The aliphatic hydrocarbon content of the SOM can give information on organisms and processes involved in SOC storage (Almendros et al. 1991; Baldock et al. 1992; Kögel-Knabner and Hatcher 1989; Nip et al. 1986). Saturated lipids and especially alkanes are more refractory to microbial oxidation than unsaturated lipids (Gocke et al. 2013). Apart from the environmental information provided by the alkane signature of the different soils, these compounds also play a role affecting different soil physical properties. Their hydrophobic character can improve soil aggregation (Dinel et al. 1991; Jambu et al. 1983, 1985) or promote soil water repellency (Jiménez-Morillo et al. 2016; Jordán et al. 2013; Rumpel et al. 2004). By means of physicochemical interactions (encapsulation, solid solution, diffusion into microporous

✉ Marco A. Jiménez-González
majimenez@mncn.csic.es

[1] Museo Nacional de Ciencias Naturales (MNCN-CSIC), c/Serrano 115-B, 28006 Madrid, Spain

[2] Universidad Autónoma de Madrid (UAM), c/Francisco Tomás y Valiente 7, 28049 Madrid, Spain

structures, hydrophobic bonding, etc.), alkanes can also condense into the structure of humic substances, accumulating in the course of SOC storage (Amblès et al. 1991; Eglinton and Logan 1991; Nebbioso et al. 2015). For these reasons, the environmental information provided by the different alkyl compounds in soil can be considered (Jansen and Wiesenberg 2017).

The main goal of this work was to contribute to a better understanding of the mechanisms of SOC sequestration in wide variety of soils differing mainly in terms of vegetation, geological substrate, and land use, using chemometric assessments that do not require the detailed knowledge of the structure of the whole system under study. A series of indices more or less responsive to SOC content were compared. Therefore, a chemical and statistical assessment was carried out, considering the alkanes released by analytical pyrolysis as indicators of the SOC content.

Materials and methods

Study area

Thirty-five soil samples at (0–10 cm) were surface collected from forest and agricultural ecosystems (Table 1) so as to reflect the wide variability in soil types. The soils differed mainly in terms of geological substrate and vegetation. After removing the litter layer, composite soil samples were prepared mixing soil material collected from three sampling points; then, soils were air-dried and sieved (< 2 mm), and all visible plant fragments were removed during sieving.

Soil physical and chemical analysis

We classified soils according to the IUSS Working Group WRB (2014) system (Table 1). Soil texture was determined by using the densimeter method (Bouyoucos 1927). Soil pH was measured in soil-water suspension (1:2.5, *w*:*w*) using an XS pH meter model pH 7 (Carpi, MO, Italy). The electrical conductivity (EC) was measured with an XS COND7 conductivity meter (Carpi, MO, Italy) using soil-water slurry (1:5, *w*:*w*). Soil bulk density was determined using a cylindrical core of known volume (Blake and Hartge 1986). The cation exchange capacity (CEC) was measured with ammonium acetate according to Juo et al. (1976). The SOC concentration was determined by wet chemical oxidation with 1 N potassium dichromate (Nelson and Sommers 1982; Walkley and Black 1934). The N was determined by micro-Kjeldahl digestion (Prince 1945). The water holding capacity (WHC) of soils was determined in the laboratory at atmospheric pressure (Klute 1986). A series of routine variables indicating the quality of the SOM were determined. In particular, different SOM fractions were isolated: free organic matter (FOM), fulvic acid (FA), humic acid (HA), and humin, in order to assess the transformation extent of the SOM. The FOM was removed by flotation with 2 M H_3PO_4 (Dabin 1976). The extractable SOM was extracted by 0.1 M $Na_4P_2O_7$ and 0.1 M NaOH (Duchaufour and Jacquin 1975). The HA and FA were separated by precipitating the alkaline soil extract with 6 M HCl. The optical density (E4) of HA was determined at 465 nm at a concentration of 0.2 mg cm^{-3} in 0.01 M NaOH (Kononova 1966). In addition, absorbance values at 616 and 572 nm in the second derivative spectrum were also measured so as to determine the concentration of fungal chromophors ascribed to 4,9-dihydroxyperylene-3,10-quinone (DHPQ) structures (Kumada and Hurst 1967; Sato and Kumada 1967). Finally, solid-state ^{13}C NMR spectra of HA were integrated to obtain semi-quantitative data for the different SOC forms. The spectra were obtained in a Bruker Avance 400 MHz, operating at a frequency of 100.63 MHz with zirconium rotors of 4 mm OD with Kel-F caps. The cross polarization (CP) was used during magic-angle spinning (MAS) of the rotor at 12.5 kHz. Between 5000 and 6000 scans were accumulated with a pulse delay of 300 ms. A ramped 1H pulse was applied during the 1-ms contact time to circumvent Hartmann-Hahn mismatches. The ^{13}C chemical shifts were calibrated using tetramethylsilane (0 ppm). The spectral ranges were chosen as already reported (González-Vila et al. 1983; Knicker 2011): alkyl + α-amino C (0–45 ppm); *N*-alkyl + OCH_3 C (45–60 ppm); *O*-alkyl C (60–110 ppm); aromatic C (110–160 ppm); carbonyl C, mainly carboxyl + amide (160–220 ppm). The ^{13}C intensity distribution was determined by integrating signal intensity over the above-mentioned chemical shift regions using MestreNova 10 software suite (Mestrelab Research).

Pyrolysis

Whole soil samples were analyzed by pyrolysis-gas chromatography-mass spectrometry (Py-GC/MS), which can determine the relative percentages of the different SOM components and does not require time-consuming extractions, fractionation, or sample pretreatments (De la Rosa et al. 2008; Jiménez-González et al. 2016).

Soil (2 g) was ground to fine powder (< 0.01 mm) using a planetary agate mill and analyzed by a PY-2020iD device (Frontier Lab Ltd., Fukushima, Japan) connected to an Agilent 6890 GC/MS system with a phenylmethylsiloxane column (Agilent HP-5MS 5%). Pyrolysis temperature was 500 °C for 1 min. Helium at a flow rate of 1 cm^3 min^{-1} was used as the carrier gas. The GC oven temperature was 50 °C for 1 min; then, it was increased to 100 °C at 30 °C min^{-1}, and from 100 to 300 °C at a rate of 10 °C min^{-1}, and kept isothermal at 300 °C for 10 min. The compounds were identified from

Table 1 Location, classification, and general characteristics of soils

Sample No.	Geographic coordinates	Soil type IUSS Working Group WRB (2014)	Texture	Vegetation
1	40° 33′ N, 4° 8′ W	Dystric Cambisol (humic)	Sandy loam	*Quercus pyrenaica*
2	41° 7′ N, 3° 34′ W	Haplic Umbrisol (hyperhumic)	Sandy loam	*Pinus sylvestris*
3	40° 23′ N, 3° 16′ W	Calcaric Cambisol (humic)	Silt loam	*Quercus ilex*
4	40° 53′ N, 3° 34′ W	Gleyic Cambisol (humic)	Sandy loam	*Fraxinus angustifolia*
5	40° 53′ N, 3° 34′ W	Dystric Cambisol (humic)	Sandy loam	*Paeonia coriacea,*
6	42° 36′ N, 3° 11′ W	Calcic Chernozem (pachic)	Sandy clay loam	Pastureland with *Micropyrum tenellum*, *Trifolium dubium*, *Trifolium campestre*, *Anacyclus clavatus*
7	40° 44′ N, 3° 42′ W	Dystric Cambisol (ochric)	Sandy loam	*Quercus rotundifolia*
8	40° 47′ N, 2° 57′ W	Leptic Kastanozems (hyperhumic)	Clay loam	*Quercus rotundifolia*
9	41° 14′ N, 3° 24′ W	Leptic Podzol (arenic)	Sandy loam	*Fagus sylvatica*
10	40° 44′ N, 3° 48′ W	Dystric Cambisol (humic)	Sandy loam	*Pinus sylvestris*
11	40°21′ N, 3° 56′ W	Dystric Cambisol (humic)	Loamy sand	*Pinus pinea*
12	40° 33′ N, 3° 43′ W	Dystric Cambisol (loamic)	Loamy sand	*Quercus rotundifolia*
13	40° 54′ N, 3°28′ W	Eutric Cambisol (humic)	Sandy loam	*Juniperus oxycedrus*
14	40° 52′ N, 3° 34′ W	Eutric Cambisol (humic)	Sandy loam	*Juniperus oxycedrus*
15	40° 58′ N, 3° 37′ W	Dystric Cambisol (humic)	Sandy loam	*Pinus pinaster*
16	40° 54′ N, 3° 53′ W	Dystric Cambisol (humic)	Sandy clay loam	*Quercus pyrenaica*
17	40° 51′ N, 3° 44′ W	Dystric Cambisol (colluvic)	Sandy loam	*Pinus sylvestris*
18	40° 45′ N, 3° 41′ W	Leptic Cambisol (humic)	Loam	*Quercus ilex*
19	43° 15′ N, 2° 51′ W	Leptic Umbrisol (loamic)	Loam	Pastureland for grazing: *Brachypodium retusum*, *Lolium perenne*, *Trifolium repens*
20	43° 4′ N, 2° 35′ W	Haplic Luvisol (humic)	Silty clay loam	*Fagus sylvatica*
21	42° 34′ N, 2° 38′ W	Eutric Cambisol (humic)	Clay loam	Pastureland for grazing: *Brachypodium retusum*, *Cynosurus cristatus*, *Trifolium repens*
22	43° 15′ N, 2° 51′ W	Haplic Umbrisol (loamic)	Loam	*Pinus radiata*
23	42° 28′ N, 8° 53′ W	Leptic Regosol (humic)	Sandy loam	*Pinus pinaster*
24	42° 36′ N, 8° 38′ W	Leptic Regosol (humic)	Sandy loam	*Pinus pinaster*
25	43° 4′ N, 8° 22′ W	Leptic Umbrisol (hyperhumic)	Loam	*Pinus pinaster*
26	28° 22′ N, 16° 39′ W	Vitric Andosol (hyperhumic)	Sandy loam	*Laurus canariensis*
27	28° 26′ N, 16° 29′ W	Leptic Regosol (arenic)	Clay loam	*Euphorbia canariensis*
28	28° 14′ N, 16° 28′ W	Leptic Regosol (arenic)	Sandy loam	Fallow: *Solanum tuberosum*
29	28° 9′ N, 16° 38′ W	Folic Umbrisol (chromic)	Sandy loam	*Pinus canariensis*
30	40° 13′ N, 4° 29′ W	Dystric Regosol (arenic)	Sand	*Pinus pinea*
31	41° 29′ N, 4° 19′ W	Eutric Cambisol (humic)	Sand	*Pinus pinea*
32	40° 18′ N, 4° 38′ W	Eutric Cambisol (arenic)	Sandy loam	*Quercus rotundifolia*
33	41° 1′ N, 3° 12′ W	Eutric Cambisol (humic)	Silt loam	*Quercus rotundifolia*
34	40° 58′ N, 3° 44′ W	Eutric Cambisol (humic)	Sandy loam	*Juniperus thurifera*
35	40° 56′ N, 3° 41′ W	Dystric Leptosol (humic)	Loam	*Juniperus thurifera*

their mass spectra acquired at 70 eV ionizing energy with an Agilent 5973 quadrupole mass spectrometer. The ion 85 trace was used to identify and quantify the alkanes. The identity of the alkanes was confirmed by their retention times (RTs), compared with those of standards, and with RTs of other pyrolytic alkyl compounds (C_{16} and C_{18} fatty acids with a diagnostic peak at m/z 60). The carbon preference index (CPI) (calculated as ratio between odd C- and even C-numbered alkanes) and the ratio of short-to-long-chain *n*-alkanes already used (Duan and He 2011) were calculated, to get insight on the origin and transformation degree of alkanes in the different environments.

Table 2 Soil physical and chemical properties

Sample No.	SOC (g kg^{-1})	C/N	Sand (g kg^{-1})	Silt (g kg^{-1})	Clay (g kg^{-1})	pH	Bulk density (g cm^{-3})	WHC (g kg^{-1})	EC (dS m^{-1})	CEC (cmol$_c$ kg^{-1})	FOM (g C·100 C g^{-1} soil)
1	41	11.3	581	261	158	5.2	0.96	619	0.488	16.4	3.5
2	67	14.8	603	320	77	3.9	0.89	652	0.385	24.4	6.6
3	96	15.3	215	758	27	7.2	0.81	814	1.143	30.8	3.4
4	87	13.3	622	264	113	6.2	0.85	894	0.747	22.0	6.9
5	48	13.1	522	352	126	5.3	0.86	657	0.418	14.4	8.1
6	17	13.9	502	240	258	7.6	1.21	507	1.025	15.9	10.7
7	18	16.0	768	156	76	5.2	1.13	428	0.290	4.5	3.9
8	87	13.3	267	415	318	6.8	0.77	833	1.326	21.5	5.4
9	32	16.4	593	318	89	5.7	1.06	511	0.490	11.6	5.2
10	140	18.1	768	116	116	5.1	0.62	1122	0.549	14.2	13.8
11	117	26.7	828	56	116	4.9	0.76	842	0.499	13.9	23.7
12	93	8.9	848	56	96	6.4	0.86	851	0.693	20.1	5.1
13	134	12.1	548	276	176	7.0	0.56	1273	1.356	19.8	14.0
14	104	18.5	648	216	136	6.5	0.67	1007	0.786	22.2	16.3
15	81	16.9	788	116	96	6.0	0.75	910	0.552	17.1	14.6
16	55	18.0	620	180	200	5.7	0.89	676	0.309	15.9	10.8
17	39	13.0	540	290	170	5.6	1.09	480	0.186	13.8	8.8
18	105	17.0	380	410	210	7.2	0.68	1206	0.767	41.9	8.5
19	41	15.8	384	374	242	4.6	0.90	624	0.381	15.8	4.7
20	44	13.4	169	537	294	5.1	0.92	704	0.491	11.3	9.1
21	57	13.9	326	291	383	6.9	0.82	943	0.645	32.8	7.0
22	27	17.0	386	403	211	4.2	0.95	613	0.324	13.3	6.0
23	133	31.0	630	210	160	3.5	0.65	1124	0.485	32.4	15.0
24	90	20.0	670	220	110	3.7	0.79	783	0.361	22.6	7.5
25	132	18.0	320	420	260	4.2	0.54	1420	0.527	31.7	8.9
26	18	22.0	610	290	100	6.8	0.98	539	0.514	28.3	0.7
27	22	12.0	420	300	280	7.7	1.06	601	0.850	18.3	12.6
28	23	14.0	600	230	170	7.3	0.99	600	0.784	29.1	9.3
29	105	27.0	680	140	180	6.4	1.01	894	0.475	20.2	12.3
30	35	20.0	935	27	38	5.7	1.11	467	0.327	4.9	17.4
31	99	25.9	913	30	58	5.7	0.64	991	0.561	9.3	23.4
32	46	16.8	717	149	135	5.4	0.94	567	0.376	12.8	5.1
33	89	23.7	378	488	134	6.2	0.64	971	0.759	17.2	23.5
34	157	21.6	637	167	196	6.6	0.59	1197	1.087	18.0	24.9
35	92	13.9	392	384	224	7.4	0.91	743	0.820	25.9	3.4

Sample No.	FA (g C·100 C g^{-1} soil)	HA (g C·100 C g^{-1} soil)	Humin (g C·100 C g^{-1} soil)	E4 (AU)	DHPQ (AU·10^4)	^{13}C NMR HA arom C (%)	^{13}C NMR HA *O*-alkyl C (%)	^{13}C NMR HA *N*-alkyl + OCH_3 C (%)	^{13}C NMR HA alkyl C (%)	^{13}C NMR HA carbonyl C (%)
1	9.9	46.0	35.4	0.78	2.48	18.9	27.4	10.8	28.4	14.6
2	10.8	47.1	28.3	1.14	1.38	27.8	18.7	8.4	32.7	12.5
3	11.0	31.5	49.7	0.66	0.82	21.9	24.2	11.3	31.4	11.3
4	15.6	35.4	38.2	0.38	0.36	16.8	24.8	11.7	35.4	11.3

Table 2 (continued)

Sample No.	FA (g C·100 C g^{-1} soil)	HA (g C·100 C g^{-1} soil)	Humin (g C·100 C g^{-1} soil)	E4 (AU)	DHPQ (AU·10^4)	^{13}C NMR HA arom C (%)	^{13}C NMR HA *O*-alkyl C (%)	^{13}C NMR HA *N*-alkyl + OCH_3 C (%)	^{13}C NMR HA alkyl C (%)	^{13}C NMR HA carbonyl C (%)
5	9.3	38.3	40.5	0.51	0.60	19.1	24.8	11.8	31.6	12.7
6	4.1	34.9	45.0	1.29	0.91	32.3	14.3	9.6	30.8	13.0
7	12.3	28.0	49.8	0.68	1.74	18.1	21.0	9.7	41.1	10.1
8	5.5	37.0	46.4	0.86	0.59	22.1	22.2	9.9	32.9	12.9
9	16.0	28.2	42.9	0.78	3.93	18.1	21.4	10.4	37.2	12.9
10	8.0	26.5	49.1	0.71	1.10	24.1	23.1	9.6	32.4	10.9
11	5.0	13.0	56.7	0.59	0.60	22.4	25.2	11.0	31.0	10.5
12	12.6	35.7	44.1	0.53	0.47	16.6	26.6	10.7	33.2	12.8
13	10.0	39.3	31.8	0.51	0.62	18.1	24.6	9.9	36.0	11.5
14	8.1	36.3	36.6	0.61	0.44	21.0	26.3	9.4	31.0	12.2
15	13.5	38.9	30.0	0.48	0.29	19.2	25.2	10.2	34.8	10.6
16	11.9	30.3	44.0	0.55	0.94	17.4	25.7	10.2	36.1	10.7
17	14.0	47.0	24.8	0.80	1.65	18.9	25.3	10.5	33.9	11.5
18	9.1	36.6	42.8	0.51	0.53	18.6	24.7	11.8	32.3	12.7
19	10.7	46.8	29.9	0.95	2.61	22.6	20.1	8.9	37.0	11.4
20	15.1	34.0	34.2	0.61	2.12	18.2	25.7	11.2	33.8	11.1
21	17.5	48.7	21.1	0.66	0.66	19.1	26.5	10.3	31.3	12.8
22	12.4	38.0	32.1	0.95	2.71	20.6	18.6	8.8	41.5	10.5
23	12.0	43.3	25.3	1.18	1.32	26.4	20.2	8.8	32.0	12.7
24	11.3	48.8	26.9	0.76	0.37	21.7	22.3	8.9	35.0	12.0
25	11.6	39.1	31.9	0.76	0.78	21.9	26.0	10.3	29.7	12.1
26	5.4	51.8	34.7	1.60	2.65	35.1	16.8	7.9	25.5	14.8
27	8.0	31.5	42.6	0.67	1.09	20.4	19.2	10.9	37.0	12.5
28	7.4	38.4	40.8	0.72	0.35	22.6	21.7	11.3	31.9	12.5
29	9.2	27.1	48.5	0.55	0.34	24.1	22.5	10.1	32.9	10.4
30	12.6	21.3	45.5	0.66	0.66	18.4	21.4	10.5	39.5	10.3
31	7.1	10.1	56.6	0.58	0.78	20.8	22.4	10.2	36.8	9.8
32	9.3	44.7	37.2	1.13	1.00	25.0	21.3	8.7	31.5	13.6
33	11.8	18.0	41.6	0.64	0.64	20.3	24.1	10.3	35.4	9.8
34	6.3	12.1	54.4	0.46	0.46	18.8	23.7	10.9	37.3	9.4
35	13.9	37.5	40.3	0.62	0.67	17.8	25.0	11.0	34.8	11.5

SOC, soil organic carbon; *WHC*, water holding capacity; *EC*, electrical conductivity; *CEC*, cation exchange capacity; *FOM*, free organic matter; *FA*, fulvic acid; *HA*, humic acid; *E4*, optical density of HA at 465 nm; *DHPQ*, peak intensity at 616 and 572 nm in the 2nd derivative visible spectra of the HAs

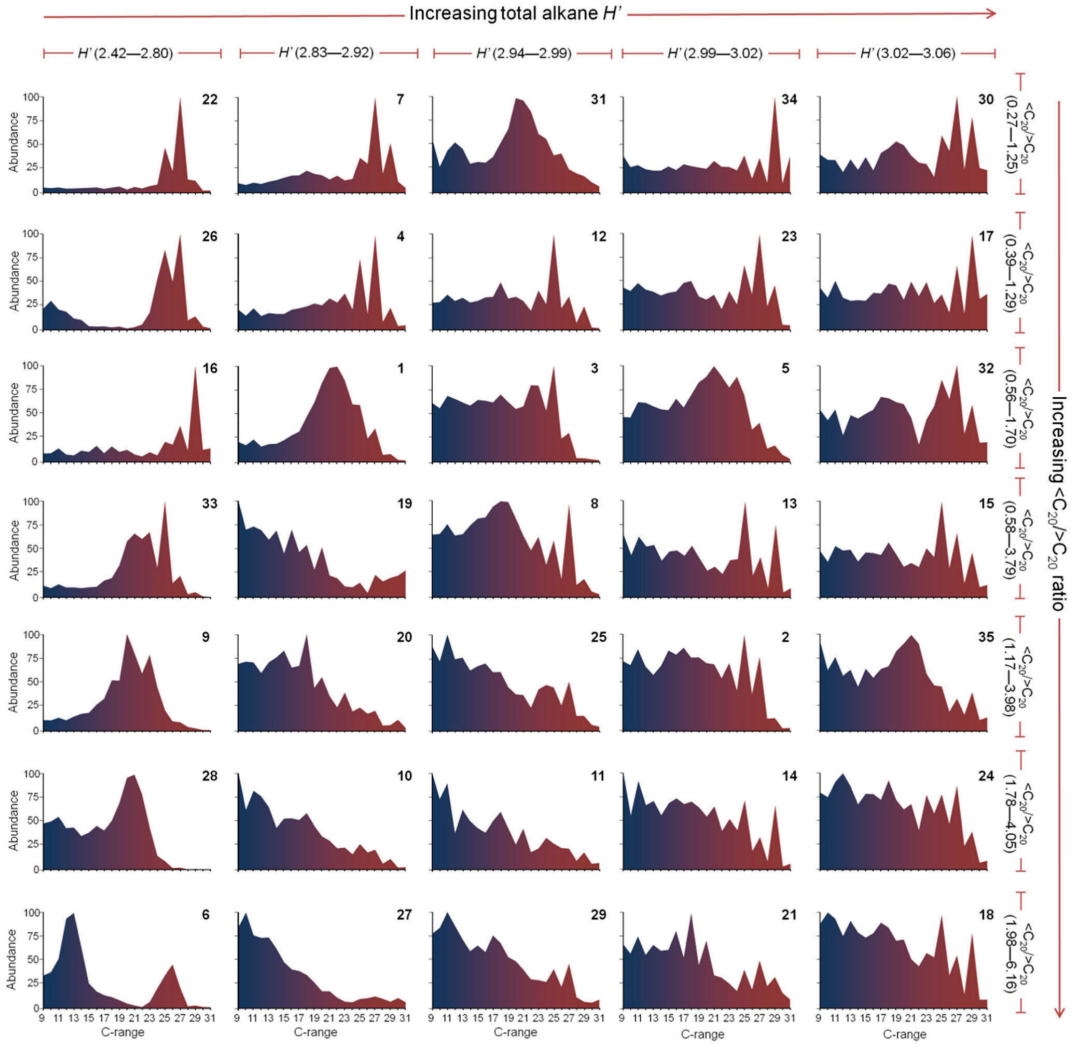


Fig. 1 Alkane homologous series released by pyrolysis from whole soil samples. Series are ordered horizontally in function of increasing diversity and vertically in function of increasing values of the $<C_{20}/>C_{20}$ ratio. Numbers of the series refer to Table 1

Statistical analysis

The variability of the concentration of SOC in the different soils (dependent variable) was described as a function of the composition of pyrolytic alkanes (independent variables) normalized as total abundances (sum of peak areas of the 23 *n*-alkanes = 100). The data were processed using either simple linear regression models or multivariate data treatments for exploring correlations between soil properties. Several indices to express diversity of the pyrolytic molecular assemblages were examined (Simpson, Margelef, and Shannon-Wiener), but the classical Shannon-Wiener diversity index (H') was selected because of its simplicity. The percentages of the 23 *n*-alkanes were used to calculate H' using the Species Diversity & Richness II software version 2.5 (Pisces Conservation). The diversity was also independently calculated for different groups of alkanes: total alkanes, short- and long-chain alkanes ($<C_{20}$ and $>C_{20}$, respectively), and odd and even C-numbered alkanes ($2n+1$ and $2n$, respectively).

The correlations between the SOC content and H' diversity indices were examined. In the second stage, principal component analysis (PCA) was carried out using the total

Table 3 Shannon H' diversity index and ratios calculated from *n*-alkanes abundances

Sample No.	$< C_{20}/> C_{20}$	CPI ($(2n+1)/2n$)	H' total alkanes	H' ($<C_{20}$)[a] alkanes	H' ($> C_{20}$)[b] alkanes	H' ($2n$)[c] alkanes	H' ($2n+1$)[d] alkanes
1	0.82	1.16	2.84	2.33	2.01	2.08	2.21
2	1.84	1.37	2.99	2.48	2.09	2.25	2.36
3	1.70	1.22	2.96	2.48	1.98	2.22	2.31
4	0.70	2.03	2.90	2.46	2.05	2.31	2.24
5	1.38	1.08	3.00	2.47	2.12	2.26	2.35
6	3.33	1.01	2.70	2.19	1.91	1.93	2.07
7	0.56	1.66	2.86	2.44	2.08	2.33	2.11
8	2.22	1.33	2.98	2.47	2.11	2.21	2.36
9	1.17	1.00	2.71	2.20	1.82	1.97	2.07
10	4.05	1.30	2.90	2.44	2.21	2.15	2.26
11	2.97	1.41	2.94	2.42	2.26	2.22	2.30
12	1.29	1.50	2.97	2.47	2.04	2.26	2.32
13	1.45	1.66	3.00	2.46	2.13	2.26	2.39
14	2.06	1.55	3.00	2.47	2.15	2.22	2.39
15	1.23	1.55	3.04	2.47	2.21	2.32	2.41
16	0.56	1.90	2.80	2.45	1.97	2.35	2.05
17	0.97	1.67	3.06	2.47	2.27	2.36	2.42
18	1.98	1.33	3.03	2.48	2.21	2.25	2.41
19	3.79	1.03	2.85	2.48	1.91	2.10	2.21
20	3.98	1.05	2.92	2.47	2.20	2.17	2.28
21	2.63	1.05	3.01	2.46	2.31	2.24	2.39
22	0.27	2.56	2.42	2.47	1.76	2.17	1.70
23	1.21	1.55	3.02	2.48	2.16	2.32	2.37
24	1.78	1.28	3.04	2.48	2.23	2.28	2.42
25	2.57	1.27	2.98	2.46	2.22	2.24	2.34
26	0.39	1.51	2.65	2.24	1.91	2.00	1.94
27	6.16	1.08	2.83	2.36	1.87	2.10	2.20
28	2.46	1.09	2.78	2.45	1.50	2.03	2.13
29	3.16	1.23	2.96	2.46	2.22	2.21	2.32
30	0.93	1.68	3.04	2.45	2.25	2.34	2.40
31	1.25	1.04	2.99	2.41	2.16	2.25	2.34
32	1.16	1.42	3.06	2.46	2.26	2.32	2.43
33	0.58	1.57	2.69	2.27	1.90	2.00	2.04
34	0.91	1.90	3.01	2.47	2.19	2.35	2.38
35	1.72	1.30	3.02	2.47	2.19	2.26	2.40

CPI, carbon preference index; H', Shannon index

[a] Short-chain: alkanes $< C_{20}$

[b] Long-chain: alkanes $> C_{20}$

[c] Even: alkanes with $2n$ C number

[d] Odd: alkanes with $2n + 1$ C number

abundances of pyrolytic alkanes and the Statistica software ver. 7.1. The PCA results were shown as a biplot where soils are displayed as points while variables (the different alkanes) are displayed as vectors in new axes represented by principal components that explain the most variability between the variables. This multivariate data treatment was used with the aim of assessing the different contributions of individual alkanes to the total variability of SOC and SOM pool content. For this purpose, and in addition to alkanes, several indicators of SOM quality, such as E4, CEC, DHPQ, H' index, and ^{13}C NMR regions of HA, were processed as supplementary variables, i.e., those which are not used to calculate the ordination axes. The principal components were also calculated only using the alkane percentages as principal variables; the supplementary

variables are represented in the space according to the components calculated from the principal variables (Legendre and Legendre 1998).

Partial least squares regression (PLS) was applied to simulate models to explain the maximum percentage of the variance of the SOC content. PLS was carried out with the Statgraphics Centurion XV software, using the total abundance of *n*-alkanes as independent variables and the SOC content as dependent variable. To select the minimum number of latent variables (LVs) of each PLS model ($P < 0.05$) the criteria considered during the cross-validation test set with the leave-one-out method were the average prediction sum of squares (PRESS) and the root mean square error (RMSE). In addition, a more rigorous criterion was used, consisting of repeating PLS using fully randomized SOC values as dependent variables (the models were discarded as overfitted if some significant ($P < 0.05$) model with random dependent variable was obtained).

Multidimensional scaling (MDS) used for variables classification was applied to explore mutual relationships between alkane abundances and other soil dependent variables, including the SOC values and using the 1-Pearson *r* correlation index as a measure of similarity. The closeness in the plane between variable scores indicates their mutual correlations. This may be used for unsupervised classification or clustering of the different variables (independent variables: alkane abundances, dependent variables: SOM descriptors). The MDS was carried out using the Statistica software version 7.1.

Results

Semi-quantitative and qualitative assessment of the SOM in the soils

The soil samples presented a large variability of SOC content (from 17 to 157 g kg^{-1}) as well as other soil properties and the SOM fractions (Table 2). As it concerns the SOM fractions (FOM, FA, and HA), they showed a large variability in amount and quality. In particular, the HA showed different values of E4 and in ^{13}C NMR spectra. This was also the case with the content of DHPQ, which may be an indicator of microbial (fungal) presence in soil (Kumada and Hurst 1967; Sato and Kumada 1967; Marynowski et al. 2015).

Alkane homologous series

The relative abundances of the alkanes, which have been obtained using the ion trace 85 *m*/*z*, are shown in Fig. 1. The series ranged between nonane (C_9) and hentriacontane (C_{31}), and generally odd C-numbered alkanes were more abundant than the even C-numbered alkanes (CPI > 1). The ratio of

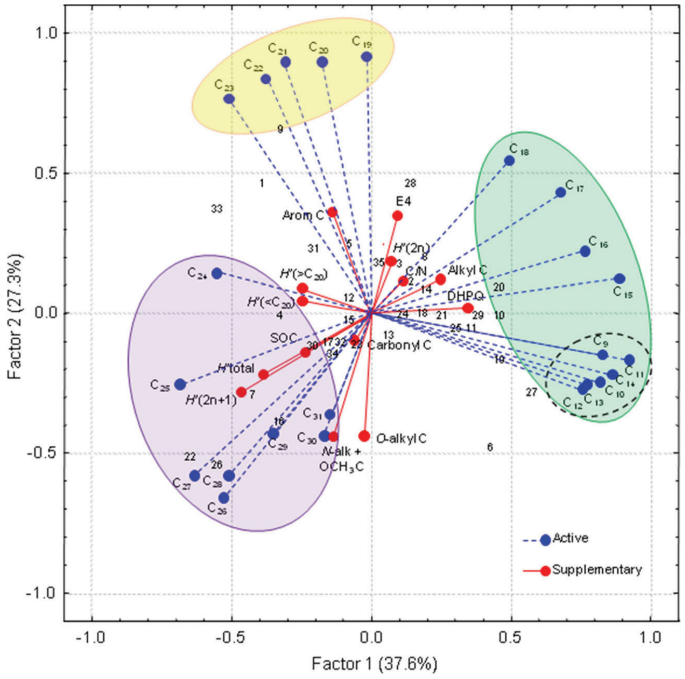


Fig. 2 Biplot obtained by principal component analysis, with numbers representing soil samples and eigenvectors pointing to the direction of increasing values of the independent variables (alkane abundances, represented with blue dashed lines). A series of SOM properties (considered as dependent variables and processed as supplementary variables) are displayed with solid red lines. The plot shows a pattern for alkanes suggesting four clusters for the independent variables. Variable labels refer to Tables 2 and 3: soil organic C (SOC), optical density of HA at 465 nm (E4), peak intensity at 616 and 572 nm in the 2nd derivative visible spectra of the HAs (DHPQ), different regions of ^{13}C NMR (arom C, alkyl C, carbonyl C, *O*-alkyl C and *N*-alk + OCH_3 C), and Shannon index (*H'*) of short-chain, long-chain, odd, and even alkanes ($< C_{20}$, $> C_{20}$, $2n + 1$ and $2n$, respectively)

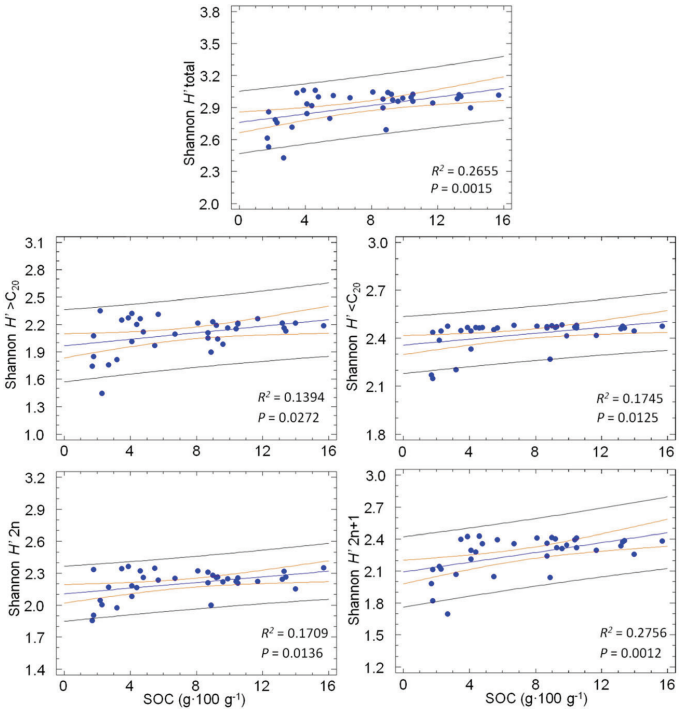


Fig. 3 Comparison of the linear relationships between the concentration of soil organic C (SOC) and the Shannon diversity index (H') calculated from total alkanes and from the short-chain ($< C_{20}$), long-chain ($> C_{20}$), even C-numbered ($2n$), and odd C-numbered ($2n + 1$) alkanes released by pyrolysis from whole soil samples. Inner (red) and outer (black) lines indicate 95% confidence and prediction limits, respectively

short-to-long-chain *n*-alkanes values is also shown in Table 3. Generally, long-chain alkanes with strong predominance of components with an odd number of C atoms (Eglinton and Hamilton 1967) are characteristic of higher plants, whereas in microbial communities and lower plants, this predominance decreases or it does not exist (Stránský et al. 1967).

Principal component analysis

Figure 2 shows the biplot obtained by PCA with the two first principal components calculated from the abundances of pyrolytic alkanes, and including a series of soil properties as supplementary variables mainly related with SOC content and quality, as well as the diversity indices obtained from the various alkane subsets. This plot explains 37.6% (first component) and 27.3% (second component) of the total variability, respectively.

Figure 2 shows that the alkane eigenvectors define three independent clusters represented by (i) short-chain alkanes (C_9–C_{18}); (ii) medium-chain alkanes from C_{19}–C_{23}; and (iii) long-chain alkanes (C_{24}–C_{31}). The short-chain alkane cluster shows a sharp independent subgroup formed by C_9–C_{14} alkanes. As regards the supplementary variables, the SOC content and the different alkane H' diversity indices were closest to each other in the plot. Therefore, diversity indices have greater correlation with SOC than the other tested supplementary variables. On the other hand, the eigenvectors of variables related to the quality of SOM are direct in the opposite direction (arom C, E4), coinciding to medium-chain length alkanes.

Linear regression models

The Shannon diversity index H' that was independently calculated for each homologous series of alkanes (Table 3) displayed a large variability within the soil samples. Whereas the diversity index H' that was calculated from the total alkane series showed a significant correlation with the SOC, different correlations were obtained for H' calculated from a narrower alkane series, i.e., only of alkanes with specific C number or chain length (Fig. 3). For instance, the H' calculated from the total alkane series showed a very significant correlation with the SOC at P value = 0.0015, $R^2 = 0.2655$, whereas the H' calculated from the odd C-numbered alkanes (P value = 0.0012, $R^2 = 0.2756$) showed better correlations than those for long- and short-chain alkanes and for even C-numbered alkanes (P value = 0.0272, $R^2 = 0.1394$; P value = 0.0125, $R^2 = 0.1745$; and P value = 0.0136, R^2 =

0.1709, respectively) although in all cases the correlations were significant at $P < 0.05$.

Partial least squares prediction models for soil C

The results obtained by PLS are in line with the previous correlations. PLS gave a highly significant ($P < 0.01$) forecasting model for predicting SOC concentration (Fig. 4a). The lack of overfitting in this model was confirmed by repeating the PLS model using randomized values of the dependent variable SOC. If the values of SOC that are not real (randomized) gave a non-significant model and the real values gave a significant one, it means that the prediction model is not based in spurious information. In all cases, with the same number of LVs (15), the random SOC values led to a poor correlation ($P > 0.05$) in the cross-validation plot (Fig. 4b). Figure 4c shows the coefficient plot and the cross-validation plot for the experimental SOC values, which correspond to the coefficients for each alkane in the equation of the prediction model generated by the PLS.

Multidimensional scaling

The results of the MDS, used to classify soil variables, are presented as a scatter diagram (Fig. 5), where the final stress was 0.2011, indicating excellent reliability of the scatterplot (Kruskal 1964). The final configuration defined by the variable scores after MDS clearly shows that the five H' diversity indices are the variables that explain to the largest extent the SOC values. Other variables (FOM, WHC, and, to lesser extent, soil C/N ratio) were also clustering close to SOC and diversity indices. On the other hand, soil properties related to SOM quality such as E4, arom C, and carbonyl C tended to form a cluster in the opposite side of the scatterplot.

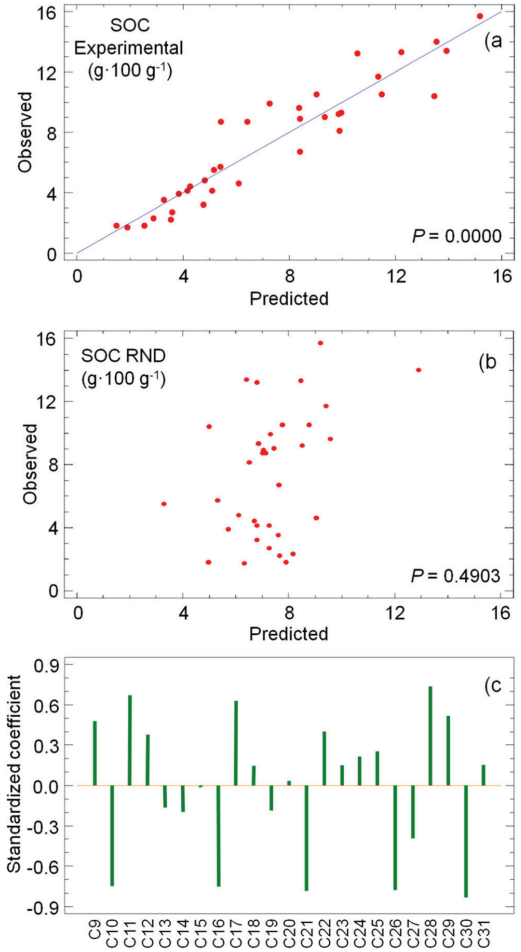


Fig. 4 Observed versus predicted values of soil organic C (SOC) content calculated by partial least squares regression (PLS) using pyrolytic alkanes from whole soil samples as descriptors to predict: **a** experimental SOC values, **b** fully randomized SOC values. Standardized coefficients of the 23 *n*-alkanes corresponding to the PLS forecasting model for the SOC are shown in **c**

Discussion

The fact that the diversity index calculated from pyrolytic *n*-alkanes showed the correlation with the SOC content may suggest a relationship between the complexity of the molecular composition of the SOM and its content in soil. This confirms previous studies, suggesting different SOM molecular structures in relation to the levels of organic C in the different soils (Almendros et al. 2017; Jiménez-González et al. 2017). The increase in alkane diversity may be related to the resistance to the biodegradation of the SOM. In soils where SOM is more effectively preserved, there are more chances of accumulation of complex molecular compounds derived from organisms inhabiting soil. Banerjee et al. (2016) suggested that the extent of SOM transformation in arable soils depends on the redundancy of bacterial and fungal communities, and Wu et al. (2010) found a significant correlation between SOC concentration and the diversity of the corresponding vegetation.

The independent PLS analysis showed that the total abundance of pyrolytic alkanes, used as independent variables, significantly predicted ($P < 0.01$) the SOC content (Fig. 4), probably because the pyrolytic alkanes accumulated information on the sources and aging of SOM. This memory of the alkane signature is probably enhanced by an increasing complexity of soil components and biogeochemical processes. In fact, pyrolytic alkanes, but not free lipid alkanes, are largely entrapped within micro-aggregates and humic-type

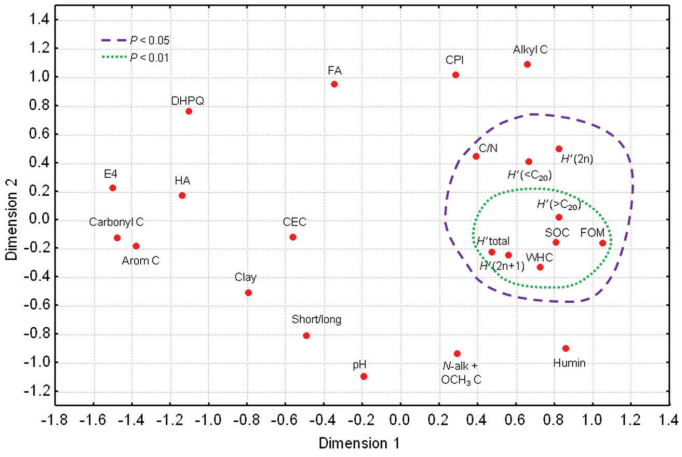


Fig. 5 Representation, using multidimensional scaling (MDS), of the relationships between different soil properties and quality descriptors of SOM, and the Shannon H' diversity index calculated from different groups of pyrolytic alkanes. Concentric lines indicate the variables positively correlated with the SOC at $P < 0.01$ (dotted line) and at $P < 0.05$ (dashed line). Variable labels refer to Tables 2 and 3

substances and thus physically protected against the degradation in soil. Therefore, pyrolytic alkanes act as a molecular record of past organisms and processes.

Different degradation techniques showed that, in the case of alkanes, pyrolysis is superior to mild wet chemical degradation in providing unbiased information (Tinoco et al. 2002), because pyrolysis caused an early thermoevaporation of alkanes, whereas chemical degradation led to continuous alteration of labile aliphatic molecules. This smoothes the alkane biogenic signature and severely modifies the indices (CPI, average chain length (ACL), etc.) that are used to define the extent of the biogeochemical transformations.

The representation of the eigenvalues of the independent variables (alkane abundances) in the space defined by the two first principal components showed different clusters. These groups can be assigned to the origin of the corresponding alkanes. Long-chain alkanes ($> C_{20}$) may derive from epicuticular waxes of vascular plants whereas short-chain alkanes ($< C_{20}$) may have a microbial origin (Simoneit and Mazurek 1982). The small cluster formed by C_9–C_{14} alkanes may derive from thermal cracking produced by the pyrolysis. In Fig. 2, the supplementary variables SOC and H' indexes (of total and odd C-numbered alkanes) are grouped close to the long-chain alkanes and this may suggest the presence of fresh SOM, as supported by the concentration of specific alkanes with the characteristic signature of waxes from higher plants. On the other hand, variables indicating SOM quality or maturity such as CEC, E4, arom C, or DHPQ tended to cluster together with medium-chain alkanes of probable microbial origin. In conclusion, despite the effectiveness of pyrolysis in breaking macromolecular structures of SOM, the alkane patterns still exhibit a characteristic biogenic signature, indicating that probably thermal cracking did not occur, and most alkanes were released in the early thermoevaporation stages.

The fact that the alkanes H' were the variables most closely related to the concentration of SOC is confirmed by the groups formed by MDS. In addition, this analysis showed that the WHC was situated near the SOC probably because retention of water in soil favors the accumulation of SOM, but because the increase in SOC content can improve soil structure and water retention (Nath 2014). The same occurred with the FOM, indicating that fresh organic matter content prevails in soils with high SOC levels. Also in the MDS analysis, indicators of SOM quality or maturity were grouped in an isolated cluster. These indicators were the E4 and the aromaticity (suggesting condensed C structures), the C=O groups, the HA content, and the concentration of DHPQ (indicative of microbial accumulation of HA-like fungal melanins), which were located in the opposed side of the plot. As a whole, the MDS plot suggests a comparatively higher quality of the humic substances in soils with higher than lower biological activity and this may lead to active mineralization of the SOM.

Conclusions

The Shannon H' diversity index of alkanes released by analytical pyrolysis shows a very significant correlation with the total SOC stored in soil. This would suggest that the diversity of alkanes may indicate the complexity of the biogeochemical processes in addition to the functional redundancy of the soil system. This makes possible the prediction of the SOC content from the signature of pyrolytic alkanes using PLS. The information supplied by the pyrolytic alkanes differs mainly in terms of their chain length, which may indicate a different origin. The long-chain alkanes (C_{24}–C_{31}), which are more correlated with fresh organic matter than other alkanes, showed the best significant correlations with the SOC levels. On the other hand, the H' indices calculated from the medium-

chain alkanes correlated better than those calculated from other alkanes with soil factors related with the SOM quality and maturity. Finally, the very short-chain alkanes (C_9–C_{18}) were unrelated to SOC content, possibly due to their origin from pyrolytic cracking of longer alkanes or condensed SOM alkyl structures.

Acknowledgements The authors are thankful to Prof. Paolo Nannipieri, Editor-in-Chief of Biology and Fertility of Soils and three anonymous referees, who have greatly contributed to improve an early version of this paper.

Funding information The study is financially supported by Spanish CICyT (grant CGL2013-43845-P). Marco A. Jiménez-González is funded by the Spanish Ministry of Economy and Competitiveness (MINECO) for his pre-doctoral FPI fellowship (BES-2014-069238).

References

Almendros G, Dorado J (1999) Molecular characteristics related to the biodegradability of humic acid preparations. Eur J Soil Sci 50:227–236. https://doi.org/10.1046/j.1365-2389.1999.00240.x

Almendros G, Sanz J, González-Vila FJ, Martín F (1991) Evidence for a polyalkyl nature of soil humin. Naturwissenschaften 78:359–362. https://doi.org/10.1007/BF01131609

Almendros G, Hernández Z, Sanz J, Rodríguez-Sánchez S, Jiménez-González MA, González-Pérez JA (2017) Graphical statistical approach to soil organic matter resilience using analytical pyrolysis data. J Chromatogr A 1533:164–173. https://doi.org/10.1016/j.chroma.2017.12.015

Amblès A, Jacquesy JC, Jambu P, Joffre J, Maggi-Churin R (1991) Polar lipid fraction in soil: a kerogen-like matter. Org Geochem 17:341–349. https://doi.org/10.1016/0146-6380(91)90097-4

Baldock JA, Oades JM, Waters AG, Peng X, Vassallo AM, Wilson MA (1992) Aspects of the chemical structure of soil organic materials as revealed by solid-state ^{13}C NMR spectroscopy. Biogeochemistry 16:1–42. https://doi.org/10.1007/BF02402261

Banerjee S, Kirkby CA, Schmutter D, Bissett A, Kirkegaard JA, Richardson AE (2016) Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. Soil Biol Biochem 97:188–198. https://doi.org/10.1016/j.soilbio.2016.03.017

Blake GR, Hartge KH (1986) Particle density. In: Klute (ed) Methods of soil analysis: part 1—physical and mineralogical methods. American Society of Agronomy and Soil Science Society of America, Madison, pp 363–375

Bouyoucos GJ (1927) The hydrometer as a new method for the mechanical analysis of soils. Soil Sci 23:343–353

Dabin B (1976) Méthode d'extraction et de fractionnement des matières humiques du sol: application à quelques études pédologiques et agronomiques dans les sols tropicaux. Cah ORSTOM ser Pédol 14:287–297

De la Rosa JM, González-Pérez JA, González-Vázquez R, Knicker H, López-Capel E, Manning DAC, González-Vila FJ (2008) Use of pyrolysis/GC-MS combined with thermal analysis to monitor C and N changes in soil organic matter from Mediterranean fire affected forest. Catena 74:296–303. https://doi.org/10.1016/j.catena.2008.03.004

Dinel H, Lévesque M, Mehuys GR (1991) Effects of long-chain aliphatic compounds on the aggregate stability of a lacustrine silty clay. Soil Sci 151:228–239. https://doi.org/10.1097/00010694-199103000-00005

Duan Y, He Y (2011) Distribution and isotopic composition of *n*-alkanes from grass, reed and tree leaves along a latitudinal gradient in China. Geochem J 45:199–207. https://doi.org/10.2343/geochemj.1.0115

Duchaufour P, Jacquin F (1975) Comparaison des processus d'humification dans les principaux types d'humus forestiers. Bull AFES 1:29–36

Eglinton G, Hamilton RJ (1967) Leaf epicuticular waxes. Science 156:1322–1335 http://www.jstor.org/stable/1721263

Eglinton G, Logan GA (1991) Molecular preservation. Philos Trans R Soc Lond 333:315–328. https://doi.org/10.1098/rstb.1991.0081

Gocke M, Kuzyakov Y, Wiesenberg GLB (2013) Differentiation of plant derived organic matter in soil, loess and rhizoliths based on *n*-alkane molecular proxies. Biogeochemistry 112:23–40. https://doi.org/10.1007/s10533-011-9659-y

González-Vila FJ, Lüdemann H-D, Martín F (1983) ^{13}C-NMR structural features of soil humic acids and their methylated, hydrolyzed and extracted derivatives. Geoderma 31:3–15. https://doi.org/10.1016/0016-7061(83)90080-0

IUSS Working Group WRB (2014) World Reference Base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.

Jambu P, Coulibaly G, Bilong P, Magnoux P, Amblès A (1983) Influence of lipids on the physical properties of soils. In: Novak B (ed) Studies about Humus. Transactions of the VIIth International Symposium. Institute of Crop Protection, Prague, 1:46–50

Jambu P, Moucawi J, Fustec E, Amblès A, Jacquesy R (1985) Interrelation entre le pH et la nature des composés lipidiques du sol: étude comparée d'une rendzine et d'un sol lessivé glossique. Agrochimica 29:186–198

Jansen B, Wiesenberg GLB (2017) Opportunities and limitations related to the application of plant-derived lipid molecular proxies in soil science. Soil 3:211–234. https://doi.org/10.5194/soil-3-211-2017

Jiménez-González MA, De la Rosa JM, Jiménez-Morillo NT, Almendros G, González-Pérez JA, Knicker H (2016) Post-fire recovery of soil organic matter in a Cambisol from typical Mediterranean forest in Southwestern Spain. Sci Total Environ 572:1414–1421. https://doi.org/10.1016/j.scitotenv.2016.02.134

Jiménez-González MA, Álvarez AM, Carral P, González-Vila FJ, Almendros G (2017) The diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of soil carbon storage. J Chromatogr A 1508:130–137. https://doi.org/10.1016/j.chroma.2017.05.068

Jiménez-Morillo NT, González-Pérez JA, Jordán A, Zavala LM, De la Rosa JM, Jiménez-González MA, González-Vila FJ (2016) Organic matter fractions controlling soil water repellency in sandy soils from the Doñana National Park (Southwestern Spain). Land Degrad Dev 27:1413–1423. https://doi.org/10.1002/ldr.2314

Jordán A, Zavala LM, Mataix-Solera J, Doerr SH (2013) Soil water repellency: origin, assessment and geomorphological consequences. Catena 108:1–5. https://doi.org/10.1016/j.catena.2013.05.005

Juo ASR, Ayanlaja SA, Ogunwale JA (1976) An evaluation of cation exchange capacity measurements for soils in the tropics. Commun Soil Sci Plant Anal 7:751–761. https://doi.org/10.1080/00103627609366684

Klute A (1986) Water retention: laboratory methods. In: Klute (ed) Methods of soil analysis: part 1—physical and mineralogical methods. American Society of Agronomy and Soil Science Society of America, Madison, pp 635–660

Knicker H (2011) Solid state CPMAS ^{13}C and ^{15}N NMR spectroscopy in organic geochemistry and how spin dynamics can either aggravate or improve spectra interpretation. Org Geochem 42:867–890. https://doi.org/10.1016/j.orggeochem.2011.06.019

Kögel-Knabner I, Hatcher PG (1989) Characterization of alkyl carbon in forest soils by CPMAS ^{13}C NMR spectroscopy and dipolar dephasing. Sci Total Environ 81–82:169–177. https://doi.org/10.1016/0048-9697(89)90122-8

Kögel-Knabner I, Zech W, Hatcher PG (1988) Chemical composition of the organic matter in forest soils: the humus layer. J Plant Nutr Soil Sci 151:331–340. https://doi.org/10.1002/jpln.19881510512

Kononova MM (1966) Soil organic matter: its nature, its role in soil formation and in soil fertility. Pergamon, Oxford

Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27. https://doi.org/10.1007/BF02289565

Kumada K, Hurst HM (1967) Green humic acid and its possible origin as a fungal metabolite. Nature 214:631–633. https://doi.org/10.1038/214631a0

Lal R (2004) Soil carbon sequestration to mitigate climate change. Geoderma 123:1–22. https://doi.org/10.1016/j.geoderma.2004.01.032

Legendre P, Legendre L (1998) Numerical ecology, vol 24, 2nd edn. Elsevier Sciences, Amsterdam

Marynowski L, Smolarek J, Hautevelle Y (2015) Perylene degradation during gradual onset of organic matter maturation. Int J Coal Geol 139:17–25. https://doi.org/10.1016/j.coal.2014.04.013

Montiel-Rozas MM, López-García A, Madejón P, Madejón E (2017) Native soil organic matter as a decisive factor to determine the arbuscular mycorrhizal fungal community structure in contaminated soils. Biol Fertil Soils 53:327–338. https://doi.org/10.1007/s00374-017-1181-5

Nath TN (2014) Soil texture and total organic matter content and its influences on soil water holding capacity of some selected tea growing soils in Sivasagar district of Assam, India. Int J Chem Sci 12:1419–1429

Nebbioso A, Vinci G, Drosos M, Spaccini R, Piccolo A (2015) Unveiling the molecular composition of the unextractable soil organic fraction (humin) by humeomics. Biol Fertil Soils 51:443–451. https://doi.org/10.1007/s00374-014-0991-y

Nelson DV, Sommers LE (1982) Total carbon, organic carbon and organic matter. In: Page AL, Miller RH, Keeney DR (eds) Methods of soil analysis: part 2, Chemical and microbiological properties, 2nd edn. American Society of Agronomy, Madison, pp 539–579

Nip M, Tegelaar EW, de Leeuw JW, Schenck PA (1986) A new non-saponifiable highly aliphatic and resistant biopolymer in plant cuticles. Naturwissenschaften 73:579–585

Pendall E, King JY (2007) Soil organic matter dynamics in grassland soils under elevated CO_2: insights from long-term incubations and stable isotopes. Soil Biol Biochem 39:2628–2639. https://doi.org/10.1016/j.soilbio.2007.05.016

Prince AL (1945) Determination of total nitrogen, ammonia, nitrates, and nitrites in soils. Soil Sci 59:47–52. https://doi.org/10.1097/00010694-194501000-00007

Requena N, Azcón M, Baca MT (1996) Chemical changes in humic substances from compost due to incubation with ligno-cellulolytic microorganisms and effects on lettuce growth. Appl Microbiol Biotechnol 45:857–863. https://doi.org/10.1007/s002530050774

Rumpel C, Seraphin A, Goebel MO, Wiesenberg G, González-Vila FJ, Bachmann J, Schwark L, Michaelis W, Mariotti A, Kögel-Knabner I (2004) Alkyl C and hydrophobicity in B and C horizons of an acid forest soil. J Plant Nutr Soil Sci 167:685–692. https://doi.org/10.1002/jpln.200421484

Sato O, Kumada K (1967) The chemical nature of the green fraction of P type humic acid. Soil Sci Plant Nutr 13:121–122. https://doi.org/10.1080/00380768.1967.10431985

Schmidt MWI, Torn MS, Abiven S, Dittmar T, Guggenberger G, Janssens IA, Kleber M, Kögel-Knabner I, Lehmann J, Manning DAC, Nannipieri P, Rasse DP, Weiner S, Trumbore SE (2011) Persistence of soil organic matter as an ecosystem property. Nature 478:49–56. https://doi.org/10.1038/nature10386

Simoneit BRT, Mazurek MA (1982) Organic matter of the troposphere—II. Natural background of biogenic lipid matter in aerosols over the rural western United States. Atmos Environ 16:2139–2159. https://doi.org/10.1016/0004-6981(82)90284-0

Stránský K, Streibl M, Herout V (1967) On natural waxes. VI. Distribution of wax hydrocarbons in plants at different evolutionary levels. Collect Czechoslov Chem Commun 32:3213–3220. https://doi.org/10.1135/cccc19673213

Tian J, Lou Y, Gao Y, Fang H, Liu S, Xu M, Blagodatskaya E, Kuzyakov Y (2017) Response of soil organic matter fractions and composition of microbial community to long-term organic and mineral fertilization. Biol Fertil Soils 53:523–532. https://doi.org/10.1007/s00374-017-1189-x

Tinoco P, Almendros G, Sanz J, González-Vila FJ (2002) Comparative analysis of the alkyl breakdown products from soil humic acids by thermal and wet chemical degradation methods. Pyrolysis 2002. 15th International Symposium of Analytical and Applied Pyrolysis. 17–20 September 2002. Leoben, Austria

Walkley A, Black IA (1934) An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Sci 37:29–37

Wu G-L, Liu Z-H, Zhang L, Hu T-M, Chen J-M (2010) Effects of artificial grassland establishment on soil nutrients and carbon properties in a black-soil-type degraded grassland. Plant Soil 333:469–479. https://doi.org/10.1007/s11104-010-0363-9

Science of the Total Environment 685 (2019) 1160–1168

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Chemometric assessment of soil organic matter storage and quality from humic acid infrared spectra

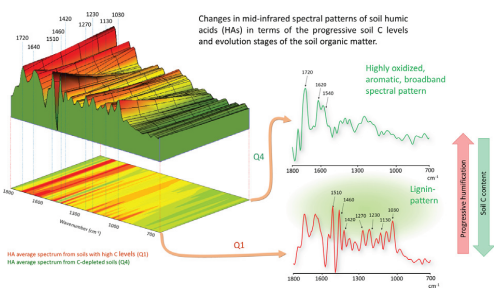
Marco A. Jiménez-González [a,b,*], Ana M. Álvarez [b], Pilar Carral [b], Gonzalo Almendros [a]

[a] Museo Nacional de Ciencias Naturales (MNCN-CSIC), c/Serrano 115-B, 28006 Madrid, Spain
[b] Universidad Autónoma de Madrid (UAM), c/Francisco Tomás y Valiente 7, 28049 Madrid, Spain

HIGHLIGHTS

- Lignin structures are major constituents of humic acids from soils of high C levels.
- Humic acids from C-depleted soils are rich in aromatic and carboxyl groups.
- Humic acid structure preserves information on soil properties as its C content.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 February 2019
Received in revised form 14 June 2019
Accepted 15 June 2019
Available online 17 June 2019

Editor: Paulo Pereira

Keywords:
Carbon sequestration
Humic substances
Optical density
Partial least squares regression
Humus quality
Nuclear magnetic resonance

ABSTRACT

The knowledge of biogeochemical mechanisms involved in soil organic carbon (SOC) storage is crucial to control its release to the atmosphere. In particular, the chemical composition of soil organic matter (SOM) plays an important role in the performance of the C storage and resilience in soils. The structural information provided by infrared spectroscopy (IR) of soil humic acid (HA) was used in the assessment of the C storage potential of 35 Spanish soils. Partial least squares (PLS) regression using the intensities of the points of the IR spectra of the HAs (4000–400 cm^{-1}) as descriptors shows that a relationship exists between IR spectral pattern and the SOC content. This was also the case for E4 (humification index based on HA optical density at 465 nm). In addition, the chemical characteristics of the HAs correlated with the SOC levels were identified from digital data treatments of the IR spectra. Additional application of principal component analysis (PCA) and multidimensional scaling (MDS) suggested that bands assigned to carboxyl and amide structures were characteristic in HAs from soils with low C content, whereas HA spectra from soils with high C levels showed a conspicuous band pattern suggesting structural units of lignin from slightly transformed plant residues. The spectral profiles were analyzed in detail by an approach based on digital subtraction of IR spectra obtained by averaging those from HAs extracted from soils in the upper and lower quartiles of the SOC distribution. The results showed that significant relationships exist between the molecular composition of HAs and SOC levels and E4 values in a way in which aromatic, carboxyl and amide groups were predominant in HAs from soils with low SOC content, whereas lignin-derived structures were more characteristic of HAs from soils with high SOC content.

* Corresponding author at: Museo Nacional de Ciencias Naturales (MNCN-CSIC), c/Serrano 115-B, 28006 Madrid, Spain.
E-mail address: majimenez@mncn.csic.es (M.A. Jiménez-González).

https://doi.org/10.1016/j.scitotenv.2019.06.231
0048-9697/

1. Introduction

Global change presumptively produced by increasing carbon dioxide (CO_2) emission to the atmosphere is a well-publicized subject of research from different scientific fields. In fact, soil represents one of the major C reservoirs at the Earth's level (Batjes, 1996; Lal, 2004), then the study of the dynamics of the soil organic carbon (SOC) and the optimization of soil management practices is interesting in terms of preventing increased emissions of CO_2 from soil (Pizzeghello et al., 2017; Hernández et al., 2019). In fact, even a small variation in the SOC pool may be reflected in the resulting atmospheric CO_2 concentration (Lal, 2004).

Biogeochemical factors responsible for SOC storage are only partially known: the stabilization of soil organic matter (SOM) by organo-mineral complexation, or physical protection including microencapsulation plays an important role in enhancing SOC sequestration (Spaccini et al., 2002; Spaccini and Piccolo, 2012; Simonetti et al., 2017). On the other hand, stabilization mechanisms depending on the chemical composition of the SOM are also very important (Solomon et al., 2007; Jiménez-González et al., 2017, 2018). In fact, SOM chemical composition depends on a plethora of biochemical and abiotic reactions such as the occasional but relevant influence of wildfires (Jiménez-González et al., 2016; Pereira et al., 2016). Most research on SOM has progressed significantly taking advantage of studies on the operationally defined humic substances (Spaccini et al., 2002; Song et al., 2014). Such humic substances can be divided into three fractions: fulvic acids (alkali- and acid-soluble fraction), humic acids (HAs) (alkali-soluble and acid-insoluble) and humin (alkali- and acid- insoluble). In particular the HA fraction, which represent one of the most structurally complex pools of the SOM, has been the subject of extensive research on its composition, reactions and hypothetical structural models (Schnitzer and Khan, 1972; Hayes et al., 1989; Hayes, 1991; Stevenson, 1994; Piccolo, 2002; Piccolo et al., 2018). In fact, recent studies take advantage of the classical idea that research on the molecular structure of humic substances represents a source of still unexplored information about biogeochemical processes (Baveye and Wander, 2019). The transformation of SOM and in particular the HA formation pathways depend on different environmental factors (Parton et al., 1987) and are expressed by means of progressive changes in the macromolecular structure of the SOM (Stevenson, 1994). In fact, the molecular composition of the HAs differs from that of the precursor plant- or microorganism-derived biomass (Miller et al., 2015; Tadini et al., 2015) and the extent of such structural differences as regards biomass could be postulated as a valid surrogate of the SOM maturity, humification degree, or SOM quality (Almendros et al., 2018; Jiménez-González et al., 2018).

Using non-destructive techniques such as infrared (IR) spectroscopy and solid-state nuclear magnetic resonance (NMR ^{13}C) it is possible to perform an exploratory analysis of soil characteristics associated with the specific HA chemical composition. Current studies have used mid-IR spectra to predict soil properties (Madari et al., 2006; Janik et al., 2007; D'Acqui et al., 2010; Cécillon et al., 2012), or to study dissolved organic matter in aquatic ecosystems (Abdulla et al., 2010). Typical bands of IR spectra have been systematically ascribed to different SOM chemical constituents and used as valuable semiquantitative descriptors of the impact of several types of environmental factors on SOM characteristics (MacCarthy and Rice, 1985; Fernández-Getino et al., 2013; Miralles et al., 2015) i.e., 1720, 1620 cm^{-1}, associated to carboxyl and aromatic groups, 2920, 1460 cm^{-1} bands for alkyl structures, 1650, 1640 cm^{-1} bands related with protein, or 1510, 1460, 1270, 1230, 1130, 1030 cm^{-1} coinciding with typical lignin structural units (Fengel and Wegener, 1984; Yang et al., 2011). Using the intensity of the above bands combined with quantitative integration data of the different regions of the NMR ^{13}C spectra of the HAs, it is possible obtain a reliable insight of the variability in the HAs' chemical constituents. Nevertheless, in the case of IR spectroscopy, further refinement based on spectral data treatments such as obtaining derivative spectra, autodeconvolution, or resolution enhancement (Almendros et al., 1992) can be very helpful for quantitative assessments, exploratory spectral pattern recognition or to improve the significance levels of chemometric models to relate spectroscopic data with different soil properties as could be SOC content or SOM quality. Although some of these treatments modify IR peak intensities, this effect is the same for homologous peaks of spectra processed by the same procedure. In particular, second derivative IR spectra are usually chosen due to it allow accurate measurement of peak intensities (Luinge et al., 1995; Viscarra Rossel, 2008).

This research includes the application of partial least squares (PLS) regression to explore soil properties reflected in the mid-IR pattern (4000–400 cm^{-1}) of soil HAs. The principal goal is to explore if there is a relationship between both the SOC content and E4 index, and the IR spectral signature of HAs isolated from the corresponding soils. In a second stage, this research aims to differentiate any specific spectroscopic pattern of HAs accumulated in soils with high potential for SOC storage, with that of HAs from C depleted soils.

2. Materials and methods

2.1. Study area

Thirty-five topsoil samples (0–10 cm) were collected from different Spanish ecosystems developed under widely differing environmental characteristics. The general characteristics of soil sampling areas were reported elsewhere (Jiménez-González et al., 2017, 2018). The soils were selected intending to cover a wide range of variability in SOC content (17–157 g C·kg $soil^{-1}$) due to explaining the variability in this soil property is a main objective of this work. Three sampling points were selected for each soil area, the samples were collected from the A horizon after removing the litter layer. Finally, composite soil samples were prepared by mixing soil subsamples from three different points, and the resulting materials were air-dried at room temperature for a week and homogenized to <2 mm.

2.2. Soil physical and chemical analysis

The soils were classified according to the World Reference Base for Soil Resources (2014) system. Soil texture was determined by the densimeter method (Bouyoucos, 1927). Soil pH was measured in soil-water suspension (1:2.5, w:w) with a pH 7 pH-meter (XS Instruments; Carpi, MO, Italy). The electrical conductivity (EC) was determined with a COND7 conductivity-meter (XS instruments; Carpi MO, Italy) using soil-water slurry (1:5, w:w). The SOC concentration was determined by wet chemical oxidation with 1 M potassium dichromate (Nelson and Sommers, 1982; Walkley and Black, 1934), and the nitrogen (N) by micro-Kjeldahl digestion (Nelson and Sommers, 1982). The water holding capacity (WHC) of the soils was measured in the laboratory at atmospheric pressure. Table 1 shows some properties of the soils and its corresponding HAs. Additional physical and chemical properties (EC, WHC, pH, soil texture, bulk density) were reported elsewhere (Jiménez-González et al., 2017, 2018)

2.3. Isolation and characterization of humic acids

The first step of the SOM fractionation consisted of removing free organic matter (FOM, particulate SOM fraction) by flotation in 2 M H_3PO_4. The humic extract was extracted with 0.1 M $Na_4P_2O_7$ and 0.1 M NaOH in reducing atmosphere. The HAs were separated from fulvic acids (FAs) by precipitating the humic extract with 6 M HCl. The HAs were purified by high-speed centrifugation (46,000 ×g) followed by dialysis (Visking tube size 6, Medicell International Ltd.) to remove soluble salts.

The optical density (E4) of the HAs was determined at 465 nm in HA solutions in 0.01 M NaOH at a concentration of 0.2 mg·cm^{-3}

Table 1
General characteristics of soil and humic acid samples.

Sample No.	SOC[a] (g·kg^{-1})	Soil C/N	E4[b] (AU)	Het-/H-Ar[c]
1	41	11.3	0.78	0.33
2	67	14.8	1.14	0.30
3	96	15.3	0.66	0.40
4	87	13.3	0.38	0.37
5	48	13.1	0.51	0.34
6	17	13.9	1.29	0.26
7	18	16.0	0.68	0.34
8	87	13.3	0.86	0.35
9	32	16.4	0.78	0.34
10	140	18.1	0.71	0.39
11	117	26.7	0.59	0.42
12	93	8.9	0.53	0.40
13	134	12.1	0.51	0.41
14	104	18.5	0.61	0.45
15	81	16.9	0.48	0.40
16	55	18.0	0.55	0.40
17	39	13.0	0.80	0.30
18	105	17.0	0.51	0.40
19	41	15.8	0.95	0.30
20	44	13.4	0.61	0.36
21	57	13.9	0.66	0.42
22	27	17.0	0.95	0.30
23	133	31.0	1.18	0.33
24	90	20.0	0.76	0.38
25	132	18.0	0.76	0.36
26	18	22.0	1.60	0.27
27	22	12.0	0.67	0.36
28	23	14.0	0.72	0.36
29	105	27.0	0.55	0.35
30	35	20.0	0.66	0.32
31	99	25.9	0.58	0.36
32	46	16.8	1.13	0.35
33	89	23.7	0.64	0.38
34	157	21.6	0.46	0.38
35	92	13.9	0.62	0.35

[a] SOC: soil organic carbon.
[b] E4: optical density of humic acid solution (0.2 mg·cm^{-3}) at 465 nm.
[c] Het-/H-Ar: ratio heteroaromatic/H-aromatic C (140–160 ppm/110–140 ppm).

(Kononova, 1982), this index is related with the humification degree and has been considered a reliable surrogate of the aromaticity of HAs (Traina et al., 1990; Tinoco et al., 2014).

The main HA structural groups were quantified using solid-state ^{13}C NMR spectroscopy using a Bruker Avance 400 MHz instrument, operating at a frequency of 100.63 MHz with 4 mm o.d. zirconium oxide rotors with Kel-F caps. The cross polarization (CP) was used during magic-angle spinning (MAS) of the rotor at 12.5 kHz. Between 5000 and 6000 scans were accumulated with a pulse delay of 300 ms. Tetramethylsilane was used to calibrate the ^{13}C chemical shifts (0 ppm). To circumvent Hartmann-Hahn mismatches, a ramped 1H-pulse was applied during the 1 ms contact time. The spectral ranges were described according to the body of classical literature (González-Vila et al., 1983; Knicker, 2011): Alkyl + α-amino C (0–45 ppm); *N*-alkyl + OCH_3 C (45–60 ppm); *O*-alkyl C (60–110 ppm); unsubstituted aromatic C (110–140 ppm); heteroaromatic C (140–160); carbonyl C (mainly carboxyl + amide, 160–220 ppm). In this study, special attention was paid to the unsubstituted aromatic C region (110–140 ppm). The ^{13}C intensity distribution was determined by integrating peak areas over the above-mentioned chemical shift regions.

2.4. Infrared spectroscopy

Fourier transform-infrared spectra were acquired with a Cary 630 FTIR spectrophotometer at a wavelength range of 4000–400 cm^{-1} and a resolution of 4 cm^{-1}. Potassium bromide (KBr) pellets containing 2 mg of powdered HA and 200 mg of KBr were scanned. Spectral data were background corrected to a reference spectrum prior to every measurement and some spurious bands, such as those from atmospheric CO_2 were removed. Resolution enhancement techniques (Almendros et al., 1992) are very useful in case of typical broadband IR spectra such as those from HAs, these spectral data treatments revealed peaks not evident in the original spectra. The method used is based on subtracting from the raw spectrum a positive multiple of its 2nd derivative (Fig. 1). Finally, 2nd derivative spectra, which show a systematic pattern of peaks and valleys and a straight virtual baseline, were chosen for semi-quantitative measurement of peak intensities (Fernández-Getino et al., 2013).

2.5. Statistical analysis

Partial least squares regression using the ParLeS software (Viscarra Rossel, 2008) was used to explain the variance of the SOC content (as dependent variable) in terms of a large series of descriptors (independent variables) consisting of the spectral points (4000–400 cm^{-1}) after mean centering. The 2nd derivative transformation of the IR spectra was selected amongst a series of standard preprocessing and pretreatment algorithms included in the program for model optimization. The PLS was applied to explore the possible relationship between the IR spectrum and SOC and E4 values, the fact that a model is possible shows that a relationship exists between them. The root mean squared error (RMSE) and the Akaike's information criterion (AIC) (Akaike, 1974) were used to select PLS models with the minimum number of factors or latent variables (LVs) i.e., to prevent overfitted spurious models. Then, the results were shown in a cross-validation plot.

In a second stage, and in order to identify spectral features characteristic of the IR spectra of HAs from soils with different SOC levels, the spectra were ordered according the C content of the corresponding soils and two subsets were considered for comparison, i.e., soils above vs. below the median of the C levels, or "soils behaving as C-sinks" vs. "soils with low SOC levels". Finally, samples were classified by quartiles (Q1 to Q4), after observing that the differences were the same but more important in the case of quartiles. The average of the IR spectra from the HAs of soils in the 1st quartile (Q1, samples with highest SOC content) and the average spectra of IR from the HAs from soils with C levels in the 4th quartile (Q4, samples with lowest SOC content) were calculated. The subtraction of the average spectra from HA samples of these quartiles was carried out to recognize the characteristics spectral bands prevailing in each subset. A similar process was carried out using the E4: the spectra were ordered according the value of E4 of the HAs, and then IR intensity values of average spectra of samples in Q1 and Q4 were subtracted.

In addition, and in an attempt to obtain more perceptual plots illustrating the spectral features of HAs from soils classified according to SOC

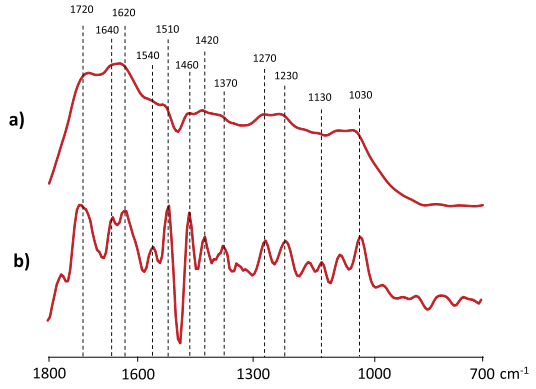


Fig. 1. Infrared spectrum of sample 34 (a) compared with the corresponding resolution enhanced spectrum (b).

or E4 values two independent spectral traces, respectively highlighting the positive or negative intensities in the above subtracted spectra were obtained. For this purpose, the average spectrum from samples in Q1 is multiplied by a factor suitable to make equal to zero the difference value at the most intense valley in the subtracted spectrum: then the resulting subtracted array (Q1–Q4) exclusively contains positive values, i.e., a virtual spectrum consisting of the characteristic bands prevailing in HA samples from the Q1 set (Manders, 1987; Martínez et al., 1990). The same operation was carried out to obtain the spectrum emphasizing the bands prevailing in HAs from soils with low SOC content, corresponding to the negative subtraction values. In this case the data array in Q1 spectrum was multiplied by a factor to obtain a spectral array where the highest value in the subtracted spectrum would coincide with the zero value, consequently all peaks in the subtraction are negative. Finally, the resulting array was multiplied by −1 to plot a positive trace. The Student *t*-test was calculated for each spectral point to illustrate the significance level ($P < 0.05$) of the spectral intensities of sample sets with different SOC levels.

The relationships between the SOC content and other soil and HA properties were examined by principal component analysis (PCA, correlation matrix, no rotation) using as variables the normalized intensities of the principal IR bands measured in the 2nd derivative and the Statistica ver. 7.1 software. This multivariate data treatment was used aiming to independently check the different contributions of individual IR bands to the variability of SOC and E4. These properties were processed as supplementary variables, i.e., those which are not used to calculate the ordination axes. The principal components were calculated using only the intensity of the selected IR bands as principal variables. Supplementary variables are represented in the space according to the components calculated from the principal variables (Legendre and Legendre, 1998).

Finally, multidimensional scaling (MDS, Kruskal, 1964) was applied to study mutual relationships between the intensity of the main IR spectral bands and soil variables, using the 1-Pearson *r* correlation index as a measure of similarity. Therefore, the proximity of the scores so calculated for the variables in the plane is considered proportional to their mutual correlations.

3. Results

3.1. ^{13}C NMR spectra

The high variability in the chemical composition of the HAs was suggested by the solid-state ^{13}C NMR spectra. The 0–45 ppm range, associated to alkyl C, was a major region of the ^{13}C NMR spectra, accounting for 25.5 to 41.5%. The signal intensity corresponding to the 45–60 ppm region (*N*-alkyl + OCH_3) showed low variation (7.9 to 11.8%). The lowest intensity in the 60–110 ppm region (*O*-alkyl C) was 14.3% and the highest value reached 27.4%. The 110–140 ppm region, corresponding to unsubstituted aromatic carbons showed values from 11.9 to 27.6% whereas the 140–160 ppm region for heteroaromatic carbons ranged between 4.3 and 7.4%. Finally, the carbonyl region, between 160 and 220 ppm showed values ranging from 9.4 to 14.8%.

3.2. Partial least squares regression

Fig. 2 shows the leave-one-out cross-validation plots for the PLS prediction model obtained for SOC (i.e., SOM quantity) and E4 (i.e., SOM quality). In the best model selected, the values of RMSE and AIC suggested including up to 6 LVs in the model for SOC prediction and 2 LVs for E4. In the cross validation plot for the PLS model using the SOC values there was a significant correlation between predicted and observed values, with $R^2 = 0.626$. In the case of E4 values, a higher correlation ($R^2 = 0.886$) than for SOC was obtained between the experimental values and the values predicted by the model.

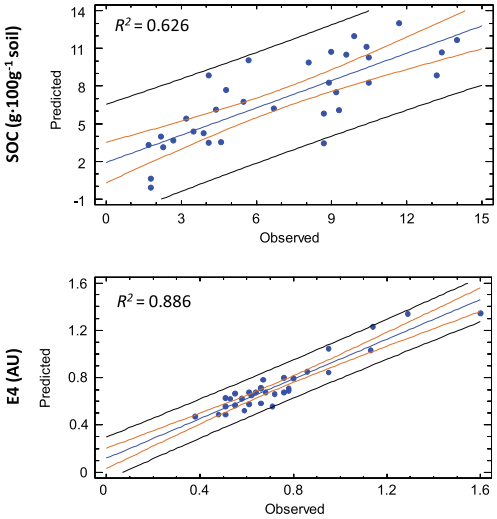


Fig. 2. Observed versus predicted values of soil organic carbon (SOC) and optical density at 465 nm of soil humic acids (E4) from Mediterranean soils calculated by partial least squares regression. Inner (orange) and outer (black) lines indicate 95% confidence and prediction limits, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

3.3. Infrared spectra

The Fig. 3 shows the 35 spectra superimposed to the corresponding resolution enhanced IR spectra. These spectra are ordered according E4 values and then according the SOC content. It is observed how the spectra with marked lignin pattern are located in the area of high SOC and low E4, whereas HAs from soils with low C levels tend to show strong carboxyl and aromatic bands.

In Fig. 4a, a comprehensive representation of the differences between spectroscopic patterns of HAs from soils with high levels of SOC as regards those with low potential for SOC storage, i.e., exposed to more severe risks of desertification, was obtained by subtracting the two spectra obtained after averaging spectra corresponding to sample sets with high and low SOC, respectively. This representation of the differential spectral features of HAs from soils in Q1 vs Q4 shows positive and negative peaks: the positive values correspond to bands predominant in the IR spectra of HAs isolated from soils with high SOC content (Q1). In the latter case, it is observed the bands 1510, 1460, 1420, 1270, 1130 and 1030 cm^{-1}. On the other side (Q4), other bands are predominant in HAs from soils with low SOC levels, i.e., bands at 1540 cm^{-1} which are related with amide groups, or 1720 cm^{-1} bands linked to C═O groups.

In the representation of the spectral profiles in which the positive and negative values are respectively extracted after zero-correction baseline by multiplying by a linear factor, the different composition of the corresponding HAs (from C-rich and C-poor soils) became much more evident. The first spectrum representing the positive region, showed the well-defined lignin pattern with sharp peaks (Fig. 4a). The second spectrum, which emphasizes the negative region of the subtraction spectrum, showed a prominent peak at 1720 cm^{-1} (C═O groups) and a broad band at 1620–1640 cm^{-1} (aromatic, unsaturated, quinone and amide structures). The same conclusion, but with opposite signs was obtained in the subtraction spectra obtained using the E4 ranks as classification factor (Fig. 4b). Negative values which are predominant in HAs with low E4 values define a spectrum with a marked lignin pattern whereas the spectrum displaying the positive values after the subtraction showed marked bands at 1720, 1620 and 1540 cm^{-1}.

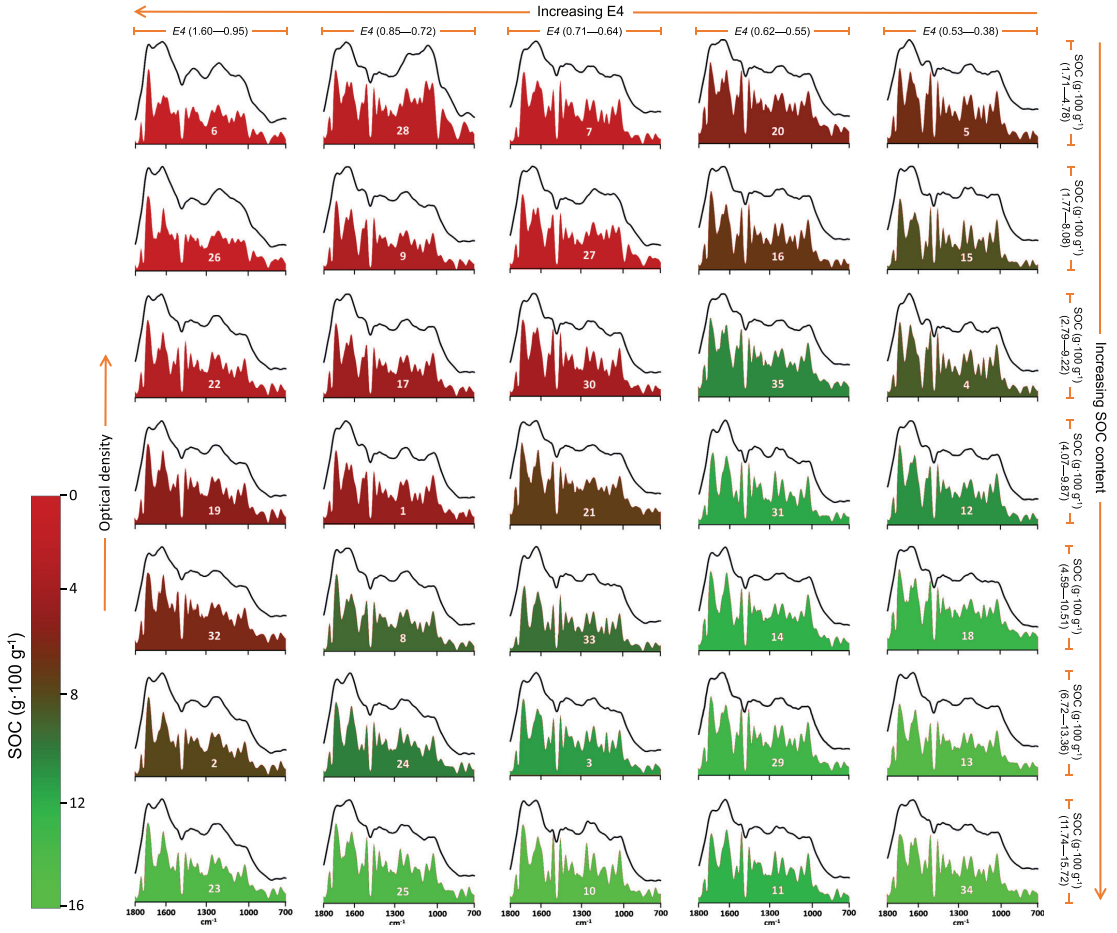


Fig. 3. Representation of infrared spectra and resolution enhanced spectra of humic acids from Mediterranean soils. Samples are ordered horizontally in function of E4 value and vertically in function of increasing content of SOC. The samples are colored according the SOC content. Numbers of the spectra refer to Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

3.4. Principal component analysis

The PCA results are shown as a biplot in Fig. 5. The two first components obtained using as variables the intensities of the principal bands of the IR spectra explained 27.4% (first component) and 19.5% (second component) of the total variance. Observing the position of the eigenvectors corresponding to the variables (band intensities of the IR spectra) used to calculate the components, it is possible to describe relationships existing between them according the location in the plane defined by the two components selected. The labels of the soil samples were located in the plane according the factors calculated. It is observed that eigenvectors for a set of bands (1510, 1460, 1420, 1370, 1270, 1130, 1030 cm^{-1}) are oriented towards a common region of the factorial space. In the same area of these bands in the plot, other soil properties like SOC, WHC, C/N ratio are located closer. On the other hand, high loading factors for bands at 1720 and 1620 cm^{-1} are defined in the opposite region near to properties such as E4, and aromatic (110–140 ppm) and carbonyl C regions of the NMR spectra (160–220 ppm).

3.5. Multidimensional scaling

The scatterdiagram in Fig. 6 shows the ordination of the variables using MDS. The final stress level in this analysis was 0.225 (Kruskal, 1964). In general, MDS analysis confirms the mutual relationships between variables suggested by PCA. In this case, the scores for variables related with SOM quality as E4, aromaticity (110–140 ppm) determined by NMR and intensity of the IR band at 1720 cm^{-1} are located closer with a significant correlation of 95% and in the opposite site to SOC level in the scatterplot. The SOC is located close to IR bands 1510, 1270, 1030 cm^{-1} and variables like WHC and Het-/H-Ar with a correlation of 95%. In addition, a set of bands (1460, 1420, 1370, 1130 cm^{-1}) and C/N are also clustering close to SOC in a correlation of 90%.

4. Discussion

The analysis by ^{13}C NMR spectroscopy showed clear differences in the chemical composition of the HAs. In particular, some ^{13}C NMR regions such as that for unsubstituted aromatic C (110–140 ppm) showed large variability. The intensity of this region is often considered to

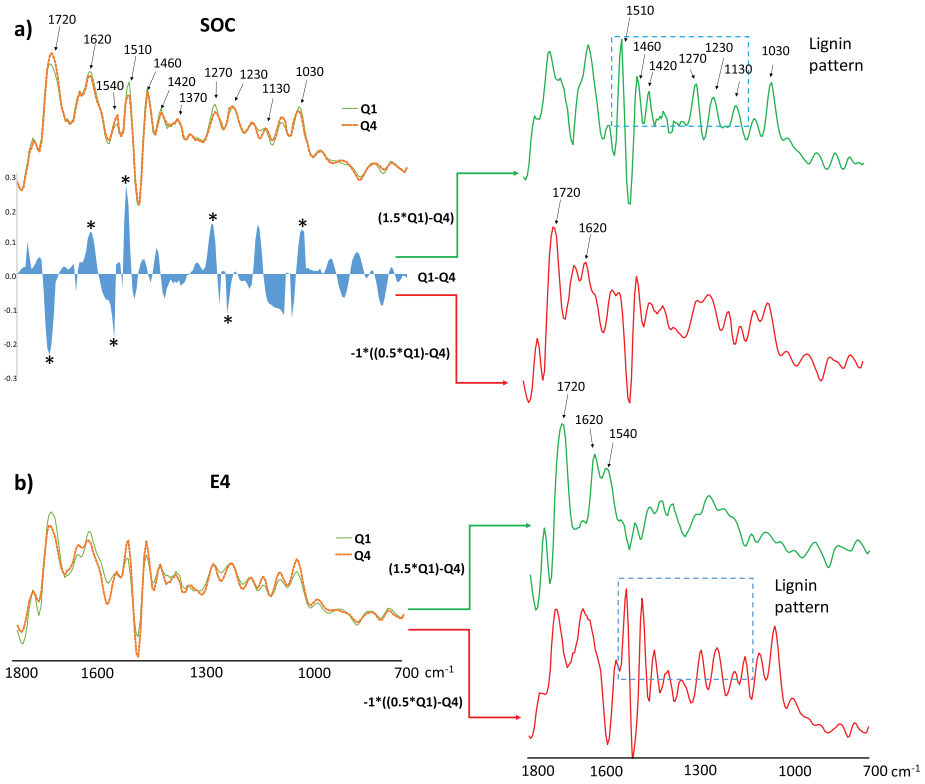


Fig. 4. Comparison of IR patterns of humic acids (HA) in terms of a) soil organic carbon (SOC) levels and b) optical density at 465 nm of soil HA (E4). The subtraction values of spectral intensities between average IR spectra from HAs in extreme quartiles of the SOC distribution in the corresponding soils is represented in a). The Q1–Q4 positive values correspond to bands predominating in C-rich soils, whereas negative values correspond to bands predominating in C-depleted soils. The asterisks (*) show values statistically different ($P < 0.05$) between band intensities of the IR spectrum in the soils with comparatively high C storage potential as regards soils comparatively poor in C (Q1 and Q4) according to the Student's *t*-test. For a more perceptual spectroscopic interpretation of the organic constituents prevailing in each of the HAs from Q1 and Q4 soil sets, two independent spectra were calculated after zero-correction baseline by multiplying by a linear factor, showing the different composition of the corresponding HAs from samples of high (green trace) and low (red trace) values of the dependent variables SOC and E4, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

parallel the degree of humification (Tinoco et al., 2014). This is due to the fact that constituents of plant and microbial biomass have a comparatively aliphatic character: both the microbial biomass and even lignin from vascular plants, have not such a high proportion of unsubstituted aromatic carbons. This result suggests great differences in the advanced stages in the humification process and could be useful to monitor the resulting maturity of the HAs.

The conspicuous prediction model obtained for SOC concentration exclusively using IR spectroscopic information indicated that a relationship exists between the functional groups of the HAs reflected in the IR spectra and the SOC level in the soils studied. Similar situation was found for E4, which was also predicted by a model exclusively using the IR spectral information. These findings suggest that the IR spectra of the HAs in the soils studied contain useful biogeochemical information about some soil characteristics such as SOC content, but also on qualitative HA properties on the transformation degree of the SOM, such as the E4, which is correlated with the proportions of carboxyl and aromatic structures (Fernández-Getino et al., 2013; Miralles et al., 2015).

The arrangement of the IR spectra in Fig. 3 suggested the accumulation of lignin-derived structures in the composition of HAs from soils with the higher SOC content and with low E4 values, whereas in soils with low SOC and high E4 this pattern is not observed. The subtraction (Fig. 4) between average IR spectra from soils belonging to quartiles in opposite levels of the SOC distribution (Q1–Q4) illustrates the predominant bands in the IR spectra according to the C storage potential of soils; the positive values (predominant bands in Q1) in the subtracted spectra reveal a characteristic IR lignin pattern viz., 1510, 1460, 1420, 1270, 1130 and 1030 cm^{-1} (Fengel and Wegener, 1984; Miralles et al., 2015). This would indicate that the soils under study with high SOC levels, display IR spectroscopic features of SOM in early transformation stages. On the other hand, negative values (predominant bands in Q4) of the subtracted spectra suggest a higher intensity of bands related with carboxyl groups (1720 cm^{-1}) and amides (1540 cm^{-1}). It is considered that these structures are characteristic of advanced humification or transformation stages (Stevenson, 1994; Jiménez-González et al., 2017, 2018; Tinoco et al., 2018). In fact, oxidized SOM with high aromaticity and N content would indicate extensive microbial reworking i.e., classically associated to an enhanced resilience of the resulting SOM (Requena et al., 1996; Almendros and Dorado, 1999; Pendall and King, 2007) driven by biogeochemical processes leading to progressively increased concentration of aromatic and carboxyl groups. This is better reflected in the two spectra illustrating—after multiplying by zero-baseline correction factors—the positive and negative peaks of the of Q1–Q4 spectral subtractions (Fig. 4). In the case of the E4 values, we found the opposite situation. Soils with high E4 not only display

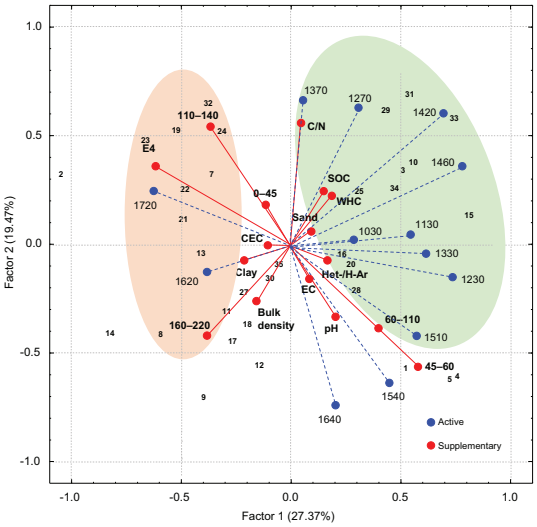


Fig. 5. Biplot obtained by principal component analysis, with numbers (1–35) representing soil samples and eigenvectors pointing to the direction of increasing values of the independent variables (IR bands, represented with blue dashed lines). A series of soil properties (also considered as dependent variables and processed as supplementary variables) are displayed with solid red lines. The biplot suggest two main clusters with variables correlated with the SOC content (encircled in green) or the aromaticity of the HA (encircled in red), respectively. Variable labels refer to Table 1. Numbers refer to peak wavenumbers in the IR spectra (1720, 1640, 1620, 1540, 1510, 1460, 1420, 1370, 1330, 1270, 1230, 1130 and 1030 cm^{-1}) and signal area in selected ranges of the ^{13}C NMR spectra (0–45, 45–60, 60–110, 110–140 and 160–220 ppm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

weak lignin pattern, but also intense 1720 cm^{-1} band, as well as prominent aromatic and amide regions (1620, 1540 cm^{-1} respectively), whereas soils with low E4 displayed well-defined IR peak lignin pattern. These facts suggest that SOM quantity and quality are linked to a different molecular composition. The HAs from soils with high SOC content tend to accumulate lignin and aliphatic structures suggesting the continuous input of fresh organic matter from vascular plants.

The biplot showing the arrangement of the studied samples and variables processed by PCA shows two groups of eigenvectors projected onto separate regions of the factorial space and presumably suggesting different C storage processes. Thus, a group of eigenvectors points to the variable corresponding to the SOC concentration, whereas others point to the main processed SOM quality descriptor, E4. In addition, eigenvectors for bands corresponding to lignin structural units (1510, 1460, 1420, 1270, 1230, 1130 cm^{-1}) are grouped in the factorial space, as could correspond to the fact that these bands conform a conspicuous pattern in the IR spectra. The intensity of these bands, which are indicating comparatively slightly transformed SOM (Miralles et al., 2015), are correlated positively with the SOC levels. On the other hand, the projections of the eigenvectors associated to WHC and SOC were very similar, probably due to high SOC levels are connected to improved soil structure and water retention (Nath, 2014; Libohova et al., 2018). The eigenvectors for E4, the intensity of the 110–140 ppm and 160–220 ppm of NMR region were oriented to the opposite side of the plot, coinciding with the scores for the samples in which the SOM evolution presumably reaches more advanced stages. In addition, the close relationship observed between these three variables and the intensities of the 1720 and 1620 cm^{-1} IR bands indicate a high functionality expected from large concentration of aromatic and *O*-containing functional groups in HAs of high degree of maturity. This coincides with the fact that the eigenvector for E4, a classical index of the degree of transformation of HAs, points to similar zones than those for the unsubstituted aromatic and carbonyl region of NMR, the 1720 cm^{-1} carboxyl and the 1620 cm^{-1} aromatic IR bands. Such facts have been interpreted as the aromatic structures and reactive *O*-containing groups represent the final products of the humification process, represented by stable structures resistant to biodegradation (Stevenson, 1994). In consequence, the contrasted IR spectroscopic features recognized in the resolution-enhanced spectra suggest inverse relationship between quality and quantity of SOC in the studied soils, i.e., soils with low SOC content would contain HAs at comparatively advanced stages of transformation, whereas soils with high SOC levels would contain HAs comparatively richer in slightly transformed and labile structures. This fact also suggests that the routine treatments used for the extraction and purification of HAs are adequate to preserve relevant chemical information

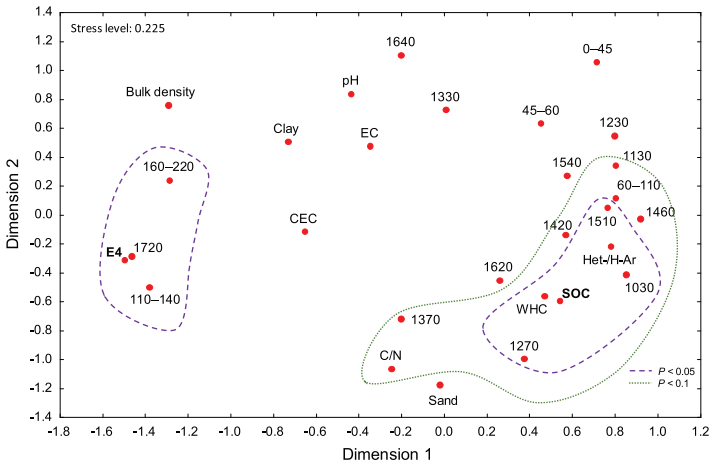


Fig. 6. Representation, using multidimensional scaling (MDS) of the relationships between different soil properties and quality descriptors of SOM, and the intensities of the main bands of the IR spectra of the corresponding soil HAs. Concentric lines indicate the variables positively correlated either with the SOC content or with the E4 at $P < 0.05$ (dashed line) and at $P < 0.1$ (dotted line). Variable labels refer to Table 1. Numbers refer to peak intensities of the IR spectra (1720, 1640, 1620, 1540, 1510, 1460, 1420, 1370, 1330, 1270, 1230, 1130 and 1030 cm^{-1}) and signal area in selected ranges of the ^{13}C NMR spectra (0–45, 45–60, 60–110, 110–140 and 160–220 ppm).

that can be recognized by IR spectroscopy and subsequently related to the origin and formation conditions of the SOM.

Finally, the MDS was useful to analyze the information about HA characteristics reflected in the IR spectra, viz., the close relationship between E4, the intensity of the 1720 cm^{-1} band and the areas of the aromatic (110–140 ppm) and carbonyl (160–220 ppm) ^{13}C NMR regions, and the negative correlation of these variables with the SOC content that was located in the opposite side of the plane (Fig. 6). Such pattern of variables supports the suggestion that high SOC levels parallel accumulation of HAs with comparatively labile and slightly transformed biomass structures.

5. Conclusions

The analysis by IR spectroscopy of HAs from a set of Mediterranean soils formed under contrasting environmental factors suggests that the organic matter of soils with different C levels also have a different molecular composition. The methodological approach suggested, based on previous validation of PLS models to explain the soil C content using the IR spectra of HA as a source of descriptors, followed by digital subtraction between average spectra of HAs from soils with extreme values of C, points out that the SOM levels paralleled the accumulation of HAs consisting of lignin-derived structures, whereas HAs in soils with low C levels present a composition in which aromatic, carbonyl and amide groups prevail. Concerning SOM quality or maturity (using E4 optical density of the HAs as comprehensive indicator) the results pointed to the opposite trend, i.e., the progressive depletion of lignin structures and accumulation of oxidized, aromatic, N-containing structures. In any case, the resolution enhanced IR spectra behaved a reliable source of additional structural information on the impact of environmental processes involved in soil C storage which are reflected in the composition of the corresponding HA fraction.

Acknowledgements

The authors would like to thank Dr. Paulo Pereira, Associate Editor of Science of the Total Environment and two anonymous reviewers, who have greatly contributed to improve an early version of this paper. Financial support by the Spanish CICYT (grant CGL2013-43845-P) is gratefully acknowledged. Marco A. Jiménez-González thanks the Spanish Ministry of Economy and Competitiveness (MINECO) for funding his pre-doctoral FPI fellowship (BES-2014-069238).

References

Abdulla, H.A.N., Minor, E.C., Dias, R.F., Hatcher, P.H., 2010. Changes in the compound classes of dissolved organic matter along an estuarine transect: a study using FTIR and ^{13}C NMR. Geochim. Cosmochim. Acta 74, 3815–3838. https://doi.org/10.1016/j.gca.2010.04.006.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Almendros, G., Dorado, J., 1999. Molecular characteristics related to the biodegradability of humic acid preparations. Eur. J. Soil Sci. 50, 227–236. https://doi.org/10.1046/j.1365-2389.1999.00240.x.

Almendros, G., González-Vila, F.J., Martín, F., Fründ, R., Lüdemann, H.-D., 1992. Solid state NMR studies of fire-induced changes in the structure of humic substances. Sci. Total Environ. 117–118, 63–74. https://doi.org/10.1016/0048-9697(92)90073-2.

Almendros, G., Hernández, Z., Sanz, J., Rodríguez-Sánchez, S., Jiménez-González, M.A., González-Pérez, J.A., 2018. Graphical statistical approach to soil organic matter resilience using analytical pyrolysis data. J. Chromatogr. A 1533, 164–173. https://doi.org/10.1016/j.chroma.2017.12.015.

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47, 151–163. https://doi.org/10.1111/j.1365-2389.1996.tb01386.x.

Baveye, P.C., Wander, M., 2019. The (bio)chemistry of soil humus and humic substances: why is the "new view" still considered novel after more than 80 years? Front. Environ. Sci. 7, 27. https://doi.org/10.3389/fenvs.2019.00027.

Bouyoucos, G.J., 1927. The hydrometer as a new method for the mechanical analysis of soils. Soil Sci. 23, 343–353.

Cécillon, L., Certini, G., Lange, H., Forte, C., Strand, L.T., 2012. Spectral fingerprinting of soil organic matter composition. Org. Geochem. 46, 127–136. https://doi.org/10.1016/j.orggeochem.2012.02.006.

D'Acqui, L.P., Pucci, A., Janik, L.J., 2010. Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. Eur. J. Soil Sci. 61, 865–876. https://doi.org/10.1111/j.1365-2389.2010.01301.x.

Fengel, D., Wegener, G., 1984. Wood: Chemistry, Ultrastructure, Reactions. Walter de Gruyter, Berlin and New York.

Fernández-Getino, A.P., Hernández, Z., Piedra Buena, A., Almendros, G., 2013. Exploratory analysis of the structural variability of forest soil humic acids based on multivariate processing of infrared spectral data. Eur. J. Soil Sci. 64, 66–79. https://doi.org/10.1111/ejss.12016.

González-Vila, F.J., Lüdemann, H.-D., Martín, F., 1983. ^{13}C-NMR structural features of soil humic acids and their methylated, hydrolyzed and extracted derivatives. Geoderma 31, 3–15. https://doi.org/10.1016/0016-7061(83)90080-0.

Hayes, M.H.B., 1991. Advances in Soil Organic Matter Research: The Impact on Agriculture and the Environment. Woodhead Publishing, Sawston.

Hayes, M.H.B., MacCarthy, P., Malcom, R.L., Swift, R.S., 1989. Humic substances II. Search of Structure. Wiley, New York.

Hernández, Z., Almendros, A., Álvarez, A.M., Figueiredo, T., Carral, P., 2019. Soil carbon stabilization pathways as reflected by the pyrolytic signature of humic acid in agricultural volcanic soils. J. Anal. Appl. Pyrolysis 137, 14–28. https://doi.org/10.1016/j.jaap.2018.10.015.

Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R., 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. Aust. J. Soil Res. 45, 73–81. https://doi.org/10.1071/SR06083.

Jiménez-González, M.A., De la Rosa, J.M., Jiménez-Morillo, N.T., Almendros, G., González-Pérez, J.A., Knicker, H., 2016. Post-fire recovery of soil organic matter in a Cambisol from typical Mediterranean forest in Southwestern Spain. Sci. Total Environ. 572, 1414–1421. https://doi.org/10.1016/j.scitotenv.2016.02.134.

Jiménez-González, M.A., Álvarez, A.M., Carral, P., González-Vila, F.J., Almendros, G., 2017. The diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of soil carbon storage. J. Chromatogr. A 1508, 130–137. https://doi.org/10.1016/j.chroma.2017.05.068.

Jiménez-González, M.A., Álvarez, A.M., Hernández, Z., Almendros, G., 2018. Soil carbon storage predicted from the diversity of pyrolytic alkanes. Biol. Fertil. Soils 54, 617–629. https://doi.org/10.1007/s00374-018-1285-6.

Knicker, H., 2011. Solid state CPMAS ^{13}C and ^{15}N NMR spectroscopy in organic geochemistry and how spin dynamics can either aggravate or improve spectra interpretation. Org. Geochem. 42, 867–890. https://doi.org/10.1016/j.orggeochem.2011.06.019.

Kononova, M.M., 1982. Materia Orgánica del Suelo: su Naturaleza, Propiedades y Métodos de Investigación. Oikos Tau, Barcelona.

Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1–27. https://doi.org/10.1007/BF02289565.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123, 1–22. https://doi.org/10.1016/j.geoderma.2004.01.032.

Legendre, P., Legendre, L., 1998. Numerical Ecology. second ed. 24. Elsevier Science.

Libohova, Z., Seyhold, C., Wysocki, D., Wills, S., Schoeneberger, P., Williams, C., Lindbo, D., Stott, D., Owens, P.R., 2018. Reevaluating the effects of soil organic matter and other properties on available water-holding capacity using the National Cooperative Soil Survey Characterization Database. J. Soil Water Conserv. 73, 411–421. https://doi.org/10.2489/jswc.73.4.411.

Luinge, H.J., van der Maas, J.H., Visser, T., 1995. Partial least squares regression as a multivariate tool for the interpretation of infrared spectra. Chemom. Intell. Lab. Syst. 28, 129–138. https://doi.org/10.1016/0169-7439(95)80045-B.

MacCarthy, P., Rice, J.A., 1985. Spectroscopic methods (other than NMR) for determining functionality in humic substances. In: Aiken, G.R., McKnight, D.M., Wershaw, R.L., MacCarthy, P. (Eds.), Humic Substances in Soils, Sediments and Water. Wiley, New York, pp. 527–559.

Madari, B.E., Reeves III, J.B., Machado, P.L.O.A., Guimarães, C.M., Torres, E., McCarthy, G.W., 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. Geoderma 136, 245–259. https://doi.org/10.1016/j.geoderma.2006.03.026.

Manders, W.F., 1987. Solid-state ^{13}C NMR determination of the syringyl/guaiacyl ratio in hardwoods. Holzforschung 41, 13–18. https://doi.org/10.1515/hfsg.1987.41.1.13.

Martínez, A.T., Barrasa, J.M., Almendros, G., González, A.E., 1990. Fungal transformation of lignocellulosics as revealed by chemical and ultrastructural analyses. In: Coughlan, M.P., Collaço, M.T. (Eds.), Advances on Biological Treatments of Lignocellulosic Materials. Elsevier, London, pp. 129–147.

Miller, K.E., Lai, C.-T., Friedman, E.S., Angenent, L.T., Lipson, D.A., 2015. Methane suppression by iron and humic acids in soils of the Arctic Coastal Plain. Soil Biol. Biochem. 83, 176–183. https://doi.org/10.1016/j.soilbio.2015.01.022.

Miralles, I., Piedra-Buena, A., Almendros, G., González-Vila, F.J., González-Pérez, J.A., 2015. Pyrolytic appraisal of the lignin signature in soil humic acids: Assessment of its usefulness as carbon sequestration marker. J. Anal. Appl. Pyrolysis 113, 107–115. https://doi.org/10.1016/j.jaap.2014.11.010.

Nath, T.N., 2014. Soil texture and total organic matter content and its influences on soil water holding capacity of some selected tea growing soils in Sivasagar district of Assam, India. Int. J. Chem. Sci. 12, 1419–1429.

Nelson, D.V., Sommers, L.E., 1982. Total carbon, organic carbon and organic matter. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis: Part 2, Chemical and Microbiological Properties, Second ed. American Society of Agronomy, Madison, WI, pp. 539–579.

Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S., 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Sci. Soc. Am. J. 51, 1173–1179. https://doi.org/10.2136/sssaj1987.03615995005100050015x.

Pendall, E., King, J.Y., 2007. Soil organic matter dynamics in grassland soils under elevated CO_2: insights from long-term incubations and stable isotopes. Soil Biol. Biochem. 39, 2628–2639. https://doi.org/10.1016/j.soilbio.2007.05.016.
Pereira, P., Rein, G., Martín, D., 2016. Past and present post-fire environments. Sci. Total Environ. 573, 1275–11277. https://doi.org/10.1016/j.scitotenv.2016.05.040.
Piccolo, A., 2002. The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science. Adv. Agron. 75, 57–134. https://doi.org/10.1016/S0065-2113(02)75003-7.
Piccolo, A., Spaccini, R., Drosos, M., Vinci, G., Cozzolino, V., 2018. The molecular composition of humus carbon: Recalcitrance and reactivity in soils. In: Garcia, C., Nannipieri, P., Hernandez, T. (Eds.), The Future of Soil Carbon. Academic Press, pp. 87–124 https://doi.org/10.1016/B978-0-12-811687-6.00004-3.
Pizzeghello, D., Francioso, O., Concheri, G., Muscolo, A., Nardi, S., 2017. Land use affects the soil C sequestration in alpine environment, NE Italy. Forests 8, 197. https://doi.org/10.3390/f8060197.
Requena, N., Azcón, M., Baca, M.T., 1996. Chemical changes in humic substances from compost due to incubation with ligno-cellulolytic microorganisms and effects on lettuce growth. Appl. Microbiol. Biotechnol. 45, 857–863. https://doi.org/10.1007/s002530050774.
Schnitzer, M., Khan, U.S., 1972. Humic Substances in the Environment. Dekker, New York.
Simonetti, G., Francioso, O., Dal Ferro, N., Nardi, S., Berti, A., Morari, F., 2017. Soil porosity in physically separated fractions and its role in SOC protection. J. Soils Sediments 17, 70–84. https://doi.org/10.1007/s11368-016-1508-0.
Solomon, D., Lehmann, J., Thies, J., Schäfer, T., Liang, B., Kinyangi, J., Neves, E., Petersen, J., Luizão, F., Skjemstad, J., 2007. Molecular signature and sources of biochemical recalcitrance of organic C in Amazonian Dark Earths. Geochim. Cosmochim. Acta 71, 2285–2298. https://doi.org/10.1016/j.gca.2007.02.014.
Song, X.-Y., Liu, S.-T., Liu, Q.-H., Zhang, W.-J., Hu, C.-G., 2014. Carbon sequestration in soil humic substances under long-term fertilization in a wheat-maize system from North China. J. Integr. Agric. 13, 562–569. https://doi.org/10.1016/S2095-3119(13)60713-3.
Spaccini, R., Piccolo, A., 2012. Carbon sequestration in soils by hydrophobic protection and in situ catalyzed photo-polymerization of soil organic matter (SOM): chemical and physical–chemical aspects of SOM in field plots. In: Piccolo, A. (Ed.), Carbon Sequestration in Agricultural Soils. Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-642-23385-2_4.
Spaccini, R., Piccolo, A., Conte, P., Haberhauer, G., Gerzabek, M.H., 2002. Increased soil organic carbon sequestration through hydrophobic protection by humic substances. Soil Biol. Biochem. 34, 1839–1851. https://doi.org/10.1016/S0038-0717(02)00197-9.
Stevenson, F.J., 1994. Humus Chemistry: Genesis, Composition, Reactions. Wiley, New York.
Tadini, A.M., Constantino, I.C., Nuzzo, A., Spaccini, R., Piccolo, A., Moreira, A.B., Bisinoti, M.C., 2015. Characterization of typical aquatic humic substances in areas of sugarcane cultivation in Brazil using tetramethylammonium hydroxide thermochemolysis. Sci. Total Environ. 518–519, 201–208. https://doi.org/10.1016/j.scitotenv.2015.02.103.
Tinoco, P., Almendros, G., González-Vila, F.J., Sanz, J., González-Pérez, J.A., 2014. Revisiting molecular characteristics responsive for the aromaticity of soil humic acids. J. Soils Sediments 15, 781–791. https://doi.org/10.1007/s11368-014-1033-y.
Tinoco, P., Almendros, G., Sanz, J., 2018. Soil perturbation in Mediterranean ecosystems reflected by differences in free-lipid biomarker assemblages. J. Agric. Food Chem. 66, 9895–9906. https://pubs.acs.org/action/showCitFormats?doi=10.1021%2Facs.jafc.8b01483.
Traina, S.J., Novak, J., Smeck, N.E., 1990. An ultraviolet absorbance method of estimating the percent aromatic carbon content of humic acids. J. Environ. Qual. 19, 151–153. https://doi.org/10.2134/jeq1990.00472425001900010023x.
Viscarra Rossel, R.P., 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemom. Intell. Lab. Syst. 90, 72–83. https://doi.org/10.1016/j.chemolab.2007.06.006.
Walkley, A., Black, I.A., 1934. An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Sci. 37, 29–37.
World Reference Base for Soil Resources, 2014. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.
Yang, Q., Wu, S., Lou, R., LV, G., 2011. Structural characterization of lignin from wheat straw. Wood Sci. Technol. 45, 419–431. https://doi.org/10.1007/s00226-010-0339-1.


CSIC
Consejo Superior de Investigaciones Científicas

mncn
museonacionaldecienciasnaturales

UA
UNIVERSIDAD AUTONOMA
DE MADRID